

PAPER • OPEN ACCESS

Transformer Fault Prediction Method Based on Multiple Linear Regression

To cite this article: Qin Jiafeng *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **486** 012037

View the [article online](#) for updates and enhancements.

Transformer Fault Prediction Method Based on Multiple Linear Regression

Qin Jiafeng¹, Zhou Chao¹, Li Longlong¹, Bai Demeng¹ and Zheng Wenjie¹

¹State Grid Shandong Electric Power Research Institute, China

Email:qinjiafeng0118@163.com

Abstract. This paper mainly analyzes the transformer fault evolution rule. Meanwhile it analyzes the correlation relation between transmission and transformation equipment fault characteristic parameter and the fault. This paper establishes the fault characteristic parameter multi-factor forecast model and the equipment fault diagnosis model, which realizes the transformer fault occurrence probability, the fault type and the fault position real-time accurate forecast. It includes data acquisition and preprocessing, multi-factor prediction of fault characteristic parameters, correlation analysis of characteristic parameters and fault types, failure probability of various faults, equipment failure probability and fault diagnosis. In this paper, a transformer characteristic parameter prediction model based on environment and other factors is proposed. The example analysis of transformer characteristic parameter multi-factor prediction model based on multiple linear regression (MRL) shows that the multi-factor prediction model can effectively consider the influence of external factors on the variation of characteristic parameter, and the prediction accuracy is higher.

1. Introduction

Power transformer is one of the most critical equipment in the power system, and the operation state of the power system is directly related to transformer's health. When the transformer fails, it may have a serious impact on the power system. If we left unchecked, it will lead to large-scale power failure and other serious social impacts. If some methods or means can be used to predict the future operation status of transformers in a timely manner and take timely and effective protective measures for the upcoming faults, the occurrence frequency of accidents can be greatly reduced. Fault diagnosis of power transformer is based on the existing online monitoring data and preventive test data. However, in the actual operation of transformers, it is necessary to predict the faults of power transformers with unstable operation and no obvious signs of faults, and judge whether they will fail in the future. If transformer faults can be predicted, various preventive measures should be taken in advance to ensure the long-term stable operation of the power grid and avoid possible economic losses.[1-3]

The algorithms applied to transformer fault prediction can be divided into the following categories:

(1) Prediction problem is essentially a "grey problem", which analyses the future trend or state according to the development law of things. Because power transformer is a complex system, there are many uncertainties between state parameters and faults. Therefore, power transformer fault prediction is a typical "grey problem".[4-5]

(2) Artificial neural network can learn from historical data and establish nonlinear model. It takes existing data as model input to predict future data. Due to this advantage, artificial neural network has been well applied in fields such as power, exploration and weather.



(3) Support vector machines (SVM) are often used for classification and have good results in regression analysis. The paper [6] proposes a new power transformer fault prediction method, this method combines support vector machine (SVM) of different kernel functions. Through comparing the forecasting accuracy of different kernel functions. It uses the combination of particle swarm optimization algorithm for kernel function of weights, and the paper uses the optimized weight regression analysis of transformer faults to take transformer fault prediction.

This paper mainly analyses the evolution law of transformer's fault and the correlation between fault characteristic parameters and faults of transformer. It establishes multi-factor prediction model of fault characteristic parameters and fault diagnosis model of equipment, which realizes real-time and accurate prediction of fault probability fault location of transmission and transformation equipment.

2. Data Preparation

The factors affecting the operation state of substation equipment are vast and complex. To be specific, all factors that have influence on the running state of power transmission and transformation equipment. Those factors can reflect the health state of power transmission and transformation equipment which should be used as evaluation indexes. In practice, not all indicators are actionable or easily quantified. Therefore, it is necessary to analyse and sort out the parameters, extracting the parameters that are easy to obtain, easy to quantify and closely related to faults. Then we set up the characteristic parameter set as the characteristic parameter. Through the collection and statistics of existing data types, the results as shown in table 1 are obtained.

Table 1 Characteristic parameters and data

Data type	Characteristic parameter
Basic information	ID, Name, Voltage classes, ...
Characteristic parameters of transformer	C2H2, CH4, C2H4, H2, ...
EMS	P, Q, I, U
Meteorological data	Temperature, Air pressure, ...

2.1. Data processing at equal intervals

Suppose the original sequence with unequal intervals: $X^{(0)} = (x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n))$. The measured time interval of each data is c . When exists $i \neq j, i, j \in \{1, 2, \dots, n-1\}, \Delta t_i \neq \Delta t_j$. Denotes that the interval of each time period is not equal, then its equal interval processing steps are shown as follows:

(1) Set the average time interval Δt_0 , and calculate the difference coefficient c of unit time period between each time period and the average time period, where $\mu(t_i) = \frac{t_i - (i-1)\Delta t_0}{\Delta t_0} \quad i \in \{1, 2, \dots, n\}$;

(2) Find the total difference of each period $\Delta x_i^{(0)}(t_i) : \Delta x_i^{(0)}(t_i) = \mu(t_i) [x_i^{(0)}(t_{i+1}) - x_i^{(0)}(t_i)]$;

(3) Calculate the grey value of equal interval points $\otimes i : \otimes i = x_i^{(0)}(t_i) - \Delta x_i^{(0)}(t_i) \quad i \in \{1, 2, \dots, n\}$;

(4) Get equal interval sequence: $\otimes X^{(0)}(t_i) = (x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n))$.

2.2. Data normalization

Different types of data have different dimensions and ranges of measured values. In order to ensure the accuracy of prediction results, it is necessary to normalize all kinds of data and change all types of data to the same range. The commonly used normalization methods are maximum and minimum method, maximum method and mean standard deviation method. This paper adopts the maximum and minimum method, which can normalize the failure prediction data to the range of $[0, 1]$. Set the sequence $x_0 = \{x_0(1), x_0(2), \dots, x_0(N)\}$, the normalized data $x_1 = \{x_1(1), x_1(2), \dots, x_1(N)\}$, and the normalization formula is as shown in the formula[7]:

$$x_1 = \{x_1(1), x_1(2), \dots, x_1(N)\} \quad (1)$$

3. Association rule data mining

A particular fault type may be directly related to only a few of characteristic parameter, while not obviously related to the others. In order to reduce the computational complexity of fault classification caused by the increase of dimension of feature quantity, the correlation degree between various fault types and each feature quantity can be obtained by using the analysis method of association rule data mining. When determining a fault, the feature quantity whose correlation degree is lower than the specified threshold will not be considered in the calculation. A high degree of correlation indicates that the fault type is closely related to the characteristic quantity. On the contrary, it indicates that this characteristic quantity is basically irrelevant to this fault type, and obviously it can improve the judgment speed and accuracy of fault diagnosis. Therefore, by quantifying the correlation between specific fault feature quantity and various fault modes, the characteristic parameters in fault diagnosis can be determined according to the degree of correlation. At present, many relational models mostly measure the change of magnitude by displacement difference (distance between points) and measure the development trend by first or second order slope[8-9]. Therefore, the correlation degree can be expressed by displacement difference and slope difference (velocity, acceleration). Let the behavior sequence of system characteristic as formula (2),

$$X_o(k) = (X_o(1), X_o(2), \dots, X_o(n)) \quad (2)$$

The behavior sequence of relevant factors is formula(3).

$$X_i(k) = (X_i(1), X_i(2), \dots, X_i(n)), i = 1, 2, \dots, m \quad (3)$$

Dengs correlation degree is a model to calculate the grey correlation degree, which fully embodies the constraints of the four axioms of grey correlation. The correlation degree of and is:

$$\gamma(X_o, X_i) = \frac{1}{n} \sum_{k=1}^n \gamma(X_o(k), X_i(k)) \quad (4)$$

$$\text{Among them: } \gamma(X_o(k), X_i(k)) = \frac{\min_i \min_k |x_o(k) - x_i(k)| + \rho \max_i \max_k |x_o(k) - x_i(k)|}{|x_o(k) - x_i(k)| + \rho \max_i \max_k |x_o(k) - x_i(k)|}$$

In the above formula, $|x_o(k) - x_i(k)|$ is called the absolute difference between X_i and X_o when time is moment k ; $\min_i \min_k |x_o(k) - x_i(k)|$ is the minimum difference between the two poles; $\max_i \max_k |x_o(k) - x_i(k)|$ is the maximum difference between the poles; ρ is the resolution coefficient, $\rho \in [0, 1]$, ρ usually is 0.5.

4. Multi-factor prediction model of transformer characteristic parameters based on multiple linear regression

The Characteristic parameters of transformers' on-line monitoring are dissolved Gas, partial discharge (PD), winding deformation, winding hot spot temperature monitoring, insulation moisture monitoring, etc. The variation of characteristic parameters with operation time is affected by many factors, such as transformer oil temperature is closely related to ambient temperature and load rate. The current and power in EMS change periodically with the change of seasons. It is of great significance to study the mutual influence of various influencing factors on the characteristic parameters and realize the multi-factor prediction of the characteristic parameters to improve the accuracy and accuracy of the prediction results and realize the transformer fault prediction.

In the regression analysis, when the causal relationship only involves the dependent variable and an independent variable, it is called the unitary regression analysis. When a causal relationship is studied involving dependent variables and two or more independent variables, it is called a multivariate regression analysis. In addition, in the regression analysis, it is divided into linear regression analysis and nonlinear regression analysis according to whether the functional expression describing the causal relationship between independent variables and dependent variables is linear or nonlinear[10-11].

In transformer and other power transmission and transformation equipment, the relationship model can be obtained by taking the equipment state score as the dependent variable, various equipment parameters as the independent variable, and using a large number of data in actual operation as the analysis sample.

For any N points $(x_k, y_k), k=1, 2, \dots, N$, they're not the same. Let $x=[x_1, x_2, \dots, x_N]^T, y=[y_1, y_2, \dots, y_N]^T$, if there is a linear relationship between x and y , then there is a linear equation of first order $y=f(x)$ to fit this data.

Let Y be the deductible value of the transformer equipment on this parameter item. X_1, X_2, \dots, X_p (rate of change, gas capacity, temperature, humidity, etc.) can be the value of each parameter, and set there is the linear relation between Y and variables X_1, X_2, \dots, X_p . So $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$, $\varepsilon \in N(0, \sigma^2)$, $\beta_0, \beta_1, \dots, \beta_p$ and σ^2 are unknown parameter.

If $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$, then the above multiple linear regression

model can be expressed in matrix form as follow formula.

$$Y = X\beta + \varepsilon \quad (5)$$

The estimated value of parameter β is $\hat{\beta}$. Next it minimizes the least squares function $Q(\beta) = (Y - X\beta)^T (Y - X\beta)$ by finding β . The least squares estimation can be proved $\hat{\beta} = (X^T X)^{-1} X^T Y$, and the regression equation can be obtained as follows.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p \quad (6)$$

This paper uses the data of four transformers in Shandong province to analyze, including active power and current. The current of main transformer 4 is used to predict the current of main transformer 1. According to the grey correlation analysis, the correlation degree between main transformer 2~4 and main transformer 1 is greater than 0.7, and the correlation degree is relatively high. 1~1000 groups of data are used as training data, and 1001-1100 groups of data are used as test data. Multiple linear regression, BPNN and KNN models are used to predict the current 1001-1200 of main transformer 1.

The error verification results of the model are shown in table 2. EMS prediction results are shown in figure 1. It can be seen from the calculation results that the mean absolute error of the three algorithms is small, and the MLR algorithm is the smallest. So this paper takes MLR as the prediction model.

Table 2 The error verification result of the three algorithms

Algorithm	MLR	BPNN	KNN
Mean absolute error	0.7725	0.9915	2.1356

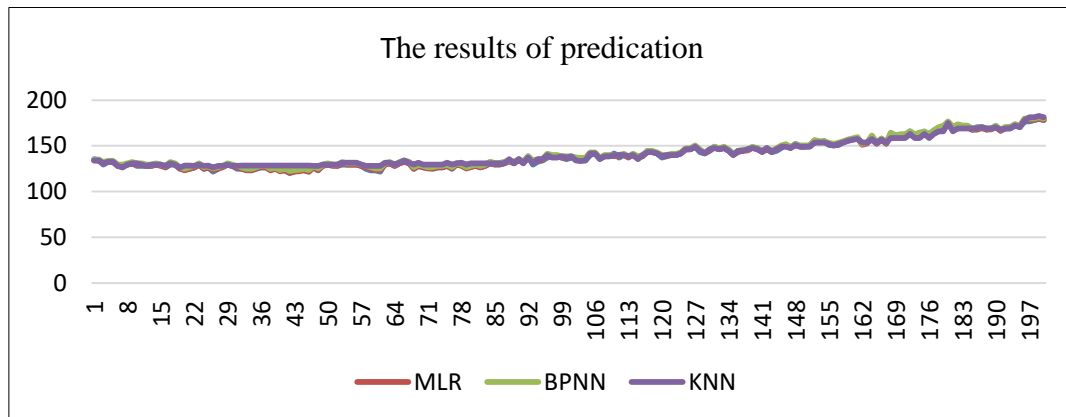


Figure 1 The prediction results of multifactor load current

Hydrogen was predicted by methane, carbon monoxide, carbon dioxide, total hydrocarbon and ethane. This paper takes groups 1-70 as the training data and groups 71-85 as the test data. Multiple linear regression, BPNN and KNN are used for prediction, and the predicted results are shown in figure 2.

It can be seen from the calculation results that when the correlation between input and output sequence is good, the prediction effect will become better as the dimension increases.

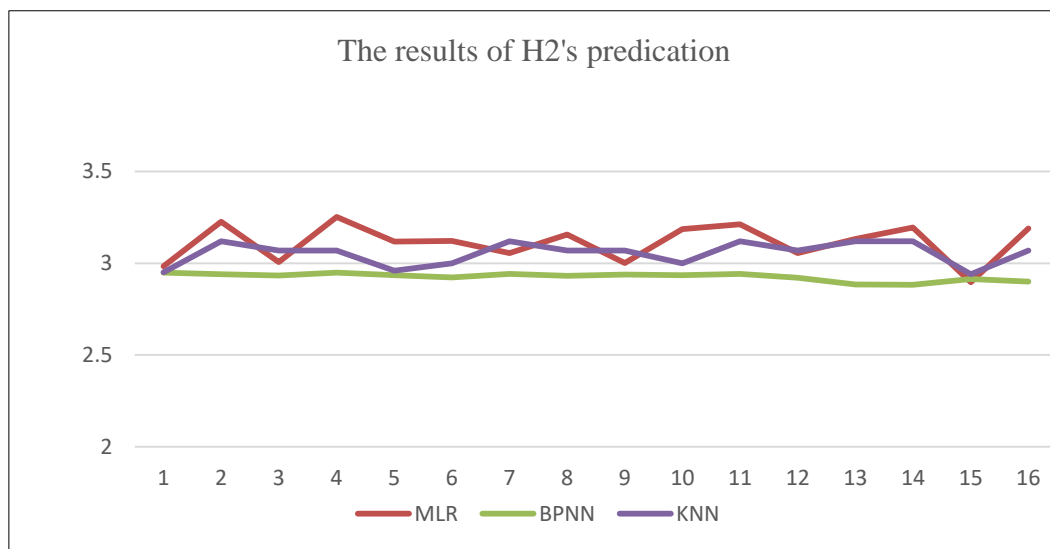


Figure2 The prediction results of H2

5. Multi-dimensional fault characteristic parameter prediction

When a certain type of fault occurs, the characteristic parameters related to the fault will change. And the probability of this type of fault should be a comprehensive function of multiple characteristic parameters. Combined with the differential threshold reflecting equipment fault, this paper proposes a comprehensive evaluation method to calculate equipment fault probability. The steps are as follows:

(1) Suppose that the device has M class fault mode; There are N fault characteristic parameters.

(2) Count the number of characteristic parameters exceeding the warning value in all fault samples $N = [n_1 \ n_2 \ \cdots \ n_N]$; The data in N are normalized, and the normalized result is the parameter weight

$$W = [\omega_1 \ \omega_2 \ \cdots \ \omega_m], \text{ where } \omega_i = \frac{n_i}{\sum_{j=1}^m n_j}.$$

(3) Calculate the probability of i class failure. As shown in formula (7):

$$P_i = \sum_{i=1}^N F_i \cdot \omega_i \quad (7)$$

(4) Equipment failure can be caused by any fault mode, so the association between equipment failure and different fault modes can be described by the reliability series model, and the calculation of equipment failure probability is shown in formula (8).

$$P = 1 - \sum_{i=1}^m 1 - P_i \quad (8)$$

Set P_f as the fault threshold. When P exceeds the threshold, the fault diagnosis shall be carried out. We calculate the failure probability of different transformers. The results are shown in table 3.

Table 3 The error verification result of the three algorithms

Data	Normal	Low energy discharge	High energy discharge	Low temperature overheating	Mid temperature overheating	High temperature overheating	Failure rate
1	0	0.15	0.42	0.44	0.54	0.48	0.406
2	0	0.14	0.45	0.4	0.55	0.46	0.4
3	0	0.09	0.39	0.40	0.58	0.48	0.388
4	0	0.08	0.42	0.39	0.61	0.50	0.4
5	0	0.07	0.43	0.38	0.63	0.52	0.406
6	0	0.07	0.45	0.38	0.64	0.54	0.416
7	0	0.07	0.46	0.37	0.66	0.55	0.422

6. Conclusion

In this paper, the association rule analysis method is adopted to analyse the transformer characteristic parameters, the correlation between faults and the evolution law, which lays a foundation for the characteristic parameters. Transformer characteristic parameter multi-factor prediction model based on multiple linear regression is proposed. Example analysis shows that the multi-factor prediction model can effectively consider the influence of external factors on the change of characteristic parameters, and the prediction accuracy is good.

References

- [1] Le Zhang. Random Forests with ensemble of feature spaces[J]. Pattern Recognition . 2014.
- [2] K. Handreck, Dabke, M.F. Conlon. An expert system for transformer impulse test diagnosis[C]. Engineering Applications of Artificial Intelligence, 1991.
- [3] A. Ratnaweera, S. K. Halgamuge, H. C. Watson. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients[C]. IEEE Transactions on Evolutionary Computation . 2004.
- [4] Kan, Hisao, Miyamoto, Teruo. Proposals for an improvement in transformer diagnosis using dissolved gas analysis (DGA) [C]. IEEE Electrical Insulation Magazine . 1995.
- [5] Bing Bing, Tiezhi Wang, Yongsheng Hao, Hongli Jia. Decision standards study on prediction technology methods[C]. Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC) . 2011.
- [6] Yao junfeng. The transformer fault prediction research of the relevance vector machine based on combined kernel function[J]. Shanxi Electric Power, 2013, 179(2):20-23.

- [7] Wang F X. Improvement on unequal interval gray forecast model[C]. Fuzzy Information and Engineering . 2006.
- [8] V. Duraisamy,N. Devarajan,D. Somasundareswari,A. Neuro fuzzy schemes for fault detection in power transformer[J] . Antony Maria Vasanth,S.N. Sivanandam. Applied Soft Computing Journal . 2006 (2).
- [9] M. H. Wang. Grey-extension method for incipient fault fore-casting of oil-immersed powertransformer[J]. Electr. Power Compon. Syst., 2004, 32(10):959-975.
- [10] Li W, Zhang Y, Guo W, et al. Forecasting Dissolved Gases Content in Power Transformer Oil Based on Particle Swarm Optimization-based RBF Neural Network[C]// Isecs International Colloquium on Computing,communication,control,and Management Proceedings. 2009:153-155.
- [11] Fei S W, Liu C L, Miao Y B. Support vector machine with genetic algorithm for forecasting of key-gas ratios in oil-immersed transformer [J]. Expert Systems with Applications An International Journal, 2009, 36(3):6326-6331.