https://doi.org/10.1051/e3sconf/202123601014

# Quantitative Prediction Method for Distribution Power Grid Risk

Tang Xuesong[1], Wu Bin[1], Yu Guangming[1], Zou Jianwei[1], Zhou Han[1], Xie Di[1], Jin Yingxu[2,*]

[1]State Grid Chongqing Electric Power Company Tongliang Power Supply Branch，Tongliang District,Chongqing , China
[2]Chongqing University of Posts and Telecommunications, Nanan District, Chongqing , China

**Abstract.** The electric power distribution grid is directly oriented to the majority of the ordinary users. Traditional operation and maintenance are performed mainly based on experience, which disable to rationally evaluate the status of the line and predict faults. Based on big data, the risk of the line is evaluated through principal component analysis in this paper, so that a machine learning algorithm is carried out to calculate the risk value of the distribution grid line unit. Finally, GA-BP neural network is used to build a line risk value prediction model for improvement.

## 1 Introduction

The traditional electric power distribution grid operation and maintenance is difficult due to a large number of distribution grid lines, which mostly relying on manual inspections. In the context of the big data era, it is extremely important to build a reasonable line failure risk prediction model using the massive data from the existing distribution grid. In literature [1], in order to extract valuable information from distribution grid data and provide effective data support for distribution grid operation, it is proposed the distribution grid risk early warning management and control based on big data. In literature [2], aiming at the problem of inadequate construction of power system facilities and measures, an optimization of BP neural network based on cloud theory and genetic algorithm is proposed to improve the accuracy of fault location. In order to solve the problem of feeding forward (BP) neural network in the cost prediction of distribution grid, a GA-BP neural network distribution network project cost prediction model is proposed [3]. In literature [4], a model for the location and isolation of the fault section in the distribution grid is proposed with dynamic topology change based on an improved genetic algorithm. In literature [5], the influence of multiple types of faults generally existing in the direct current power grid is analyzed, and a real-time fault diagnosis method based on the neural network model of current are proposed. In this paper, big data is introduced so as to build a risk value prediction model using GA-BP neural network so that evaluate the risk of the distribution grid lines through calculating the risk value of distribution grid lines based on the machine learning.

## 2 Calculate risk value of the route

According to the collected data of a power supply company's 10kV distribution line failures in two years, analysis and cause classification are carried out, so that 8 key factors leading to the distribution grid line failures are achieved.

### 2.1. Failure cause factors description

The main fault cause factors are show in **Figure 1,** with corresponding explanation.

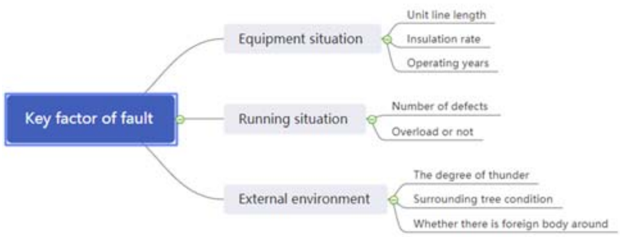



**Figure 1.** Fault cause factor and its explanation

### 2.2. Principal component analysis method

Taking into account the difference between the dimensions of the data of each variable and the inconsistent units, the method of normalizing variables is adopted to eliminate the influence of the dimension. The normalization formula is (1):

$$X_{norm} = \frac{X - X_{MIN}}{X_{MAX} - X_{MIN}} \tag{1}$$

After the data is normalized, the principal component analysis is performed to obtain the contribution rate of each principal component in **Table 1**.

[a] Corresponding author: longhy@cqupt.edu.cn

**Table 1** The contribution rate of each principal component

| | COMP. 1 | COMP. 2 | COMP. 3 | COMP. 4 | COMP. 5 | COMP. 6 | COMP. 7 | COMP. 8 |
|---|---|---|---|---|---|---|---|---|
| **STANDARD DEVIATION** | 1.4155 | 1.1296 | 1.0415 | 1.0222 | 0.9365 | 0.8680 | 0.8105 | 0.5507 |
| **PROPORTION OF VARIANCE** | 0.2504 | 0.1595 | 0.1356 | 0.1306 | 0.1096 | 0.0942 | 0.0821 | 0.0379 |
| **CUMULATIVE PROPORTION** | 0.2504 | 0.4099 | 0.5455 | 0.6762 | 0.7858 | 0.8800 | 0.9621 | 1 |

As show in **Table 1,** the cumulative contribution rate of the first 7 principal components has reached 96.21%, so that only the first 7 principal components are selected for subsequent calculations, in order to reduce the calculation dimension.

In order to obtain the comprehensive score of the principal component of each route, the value of each principal component is used to measure the magnitude of the risk value. The contribution rate of each principal component is set to construct a comprehensive evaluation function to calculate the risk score of each route in (2).

$$score = \sum_{i=1}^{7} \alpha_i Z_i \quad (2)$$

**Table 2** Comprehensive principal component scores of 3 partial routes

| Line unit name | score |
|---|---|
| **10kV Liangba Line925（2segment）** | 1.4804 |
| **10kV Liangba Line925（1segment）** | 0.9681 |
| **10kV Liangban Line915** | 2.4074 |
| **10kV Liangping line921** | 2.3284 |
| **10kV Liangtai line 911** | 1.8476 |
| **10kV Jijiu line913** | 0.6900 |
| **10kV Cuiying 6 branch line of the old line** | -0.3679 |
| **10kV Jijiu Line Iron and Stone 5 Branch Line** | -0.4243 |
| **10kV Jijiu line and fish splashes 1 branch line** | -0.1802 |
| **10kV Liangba Line Denggao 3 Branch Branch** | -0.3827 |

Since each variable in the principal component analysis is a positive indicator. That is, the larger the variable value, the greater the risk of the route. Therefore, as for the line risk value calculated through the above steps, it can also be considered that the larger the value, the greater the risk of the line.

The risk score calculated by the principal component analysis method has a negative value, which is transformed to obtain the risk value $risk\_value$, with the following formula and the actual value of the dependent variable in the regression equation (3).

$$risk\_value = [score + abs(\min(score))] \times 10 \quad (3)$$

## 3 GA-BP neural network for the risk model

### 3.1. A risk model with BP neural network

The aforementioned EMLR model has a good effect on a specific data set, while does not work well on other data sets. Thus, a BP neural network is performed to improve the model so that it can be better adapted for the data sets.

The design of BP neural network mainly includes several aspects such as the number of network layers, the number of input layer nodes, the number of hidden layer nodes, the number of output layer nodes, the transfer function, the training method, and the setting of training parameters.

(1) Number of network layers: BP neural network can contain one or more hidden layers. However, theoretically it has been proved that a single hidden layer network can result in nonlinear mapping by appropriately increasing the number of neuron nodes. Therefore, for most application, a single hidden layer is rational.

(2) Number of nodes in the input layer: The number of nodes in the input layer depends on the dimension of the input vector.

(3) Number of hidden layer nodes: The number of hidden layer nodes has a great influence on the performance of BP neural network. Generally speaking, more hidden layer nodes can bring better performance, but may lead to too long training time, which is also a defect of the current BP neural network. At present, the main empirical formulas for determining the number of hidden layer nodes are in (4), (5) :

$$M = \sqrt{n+m} + a \quad (4)$$

$$M = \log_2 n \quad (5)$$

Where, $m$ and $n$ are the number of neurons in the output layer and input layer respectively, $a \in [0,10]$。In this model, 8 key factors are used as the input of the BP neural network. The risk value is taken as the output of the BP neural network and set the number of hidden layer nodes $M$=3.


**Figure 2.** BP neural network structure diagram

80% of the samples in the data set are randomly selected as the training set so as to train the BP neural network. 20% of the samples are selected as the test set in order to test the training effect of the BP neural network. The partial result of the test set prediction is in **Table 3** as follows :

https://doi.org/10.1051/e3sconf/202123601014

**Table 3** .Comparison of prediction results of BP neural network test set with real values

| Line name | actual value | Predictive value |
|---|---|---|
| 10kV liangba line 925（2segment） | 25.3389 | 25.5762 |
| 10kV Jijiu line 913 | 14.7275 | 14.7037 |
| 10kV Jijiu lineand fish splashes 1 branch line | 9.3950 | 9.3904 |
| 10kV Liangba Line Gaotan 8 Society #2 Branch Line | 10.1373 | 10.1293 |
| 10kV Liangba Line Three Towers 3 Branch Branch | 11.1666 | 11.1630 |
| 10kV Liangban Line Baita 5 Branch Branch | 6.9587 | 6.9832 |
| 10kV Liangban Line Banzhu Street #1 Branch Line | 6.9309 | 6.9571 |
| 10kV Liangban Line Benqing No. 9 Branch Line | 8.7560 | 8.7621 |
| 10kV Liangban Line Longtai Branch 3 Branch Line | 11.7816 | 11.7846 |
| 10kV Liangban Line Longtaishe Branch Line | 10.2233 | 10.2177 |

In order to quantitatively evaluate the prediction effect of the BP neural network, the coefficient of determination is introduced, which ranges within [0, 1]. The closer it is to 1, the better the prediction effect of the model is. On the contrary, if it is closer to 0, it seems that the prediction effect of the model is worse. The expression of the coefficient of determination is in (4):

$$R^2 = \frac{(l\sum_{i=1}^{l}\hat{y}_i y_l - \sum_{i=1}^{l}\hat{y}_i\sum_{i=1}^{l}y_i)^2}{(l\sum_{i=1}^{l}\hat{y}_i^{\,2} - (\sum_{i=1}^{l}\hat{y}_i)^2)(l\sum_{i=1}^{l}y_i^2 - (\sum_{i=1}^{l}y_i)^2)} \quad (4)$$

Where, $l$ is the number of samples, $y_i\,(i = 1, 2, ..., l)$ is the true value of the i sample, and $\hat{y}_i\,(i = 1, 2, ..., l)$ is the predicted value of the *i*th sample.

## 3.2.Optimal BP neural network through the genetic algorithm

If only the traditional BP neural network is used to train the risk value of each line, the training model obtained will not be stable. In order to increase the reliability of the model, the genetic algorithm is introduced to optimize the initial weights and thresholds of neurons in the hidden layer and output layer in the BP neural network.

When using the genetic algorithm to update the weights and thresholds of each neuron, the fitness function is defined as (5):

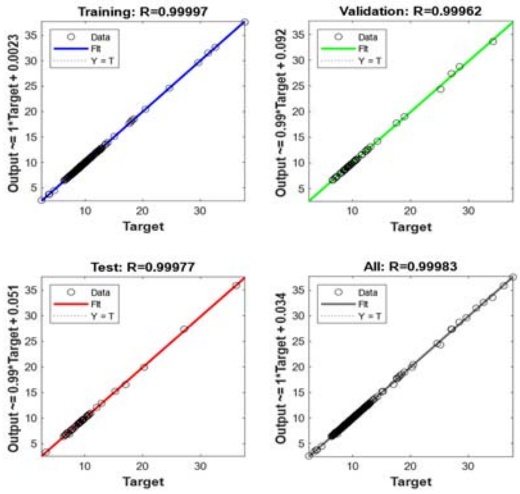
$$F = k(\sum_{i=1}^{m} abs(y_i - o_i)) \quad (5)$$

Where, $k$ is the coefficient, $m$ is the number of output nodes of the network, $y_i$ is the expected output of the *i*th node of the BP neural network, $o_i$ is the predicted output of the *i*th node.

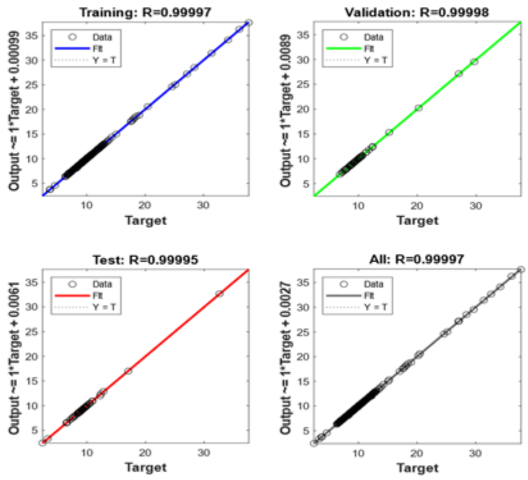
Some of the results of the test set prediction are as follows.The coefficient of determination $R^2 = 0.9999$， is much better than the effect of the BP network only used for training.

**Table4** Comparison of GA-BP algorithm test set prediction results and real values

| Line name | Actual value | Predictive value |
|---|---|---|
| 10kV Lingba line 925（2segment） | 25.3389 | 25.4497 |
| 10kV JiJiu line913 | 14.7275 | 14.8305 |
| 10kV The Jijiu line and fish splashes 1 branch line | 9.3950 | 9.3872 |
| 10kV Liangba Line Gaotan 8 Society #2 Branch Line | 10.1373 | 10.1332 |
| 10kV Liangba Line Three Towers 3 Branch Branch | 11.1666 | 11.1590 |
| 10kV Liangban Line Baita 5 Branch Branch | 6.9587 | 6.9698 |
| 10kV Liangban Line Banzhu Street #1 Branch Line | 6.9309 | 6.9416 |
| 10kV Liangban Line Benqing No. 9 Branch Line | 8.7560 | 8.7467 |
| 10kV Liangban Line Longtai Branch 3 Branch Line | 11.7816 | 11.7697 |
| 10kV Liangban Line Longtaishe Branch Line | 10.2233 | 10.2293 |




(A)




(B)

**Figure 3.** Comparison of training effects between BP (A) and GA-BP (B)

## 4 Conclusions

This article firstly conducts principal component analysis on the 8 factors that affect line faults, and obtains the principal component scores of each line, which is regarded as the risk worthy of the line. Then, 8 failure factors are taken as independent variables, and the risk value normalized by risk value score are used as the dependent variable. In addition, stochastic gradient descent method is performed for parameter training in multiple linear and exponential regression equations, so that achieve the quantitative expression of the risk value. Finally, in order to increase the generalizability of the model, 8 fault factors and risk values are performed with GA-BP network training to obtain a universal line risk value prediction model.

## Acknowledgments

This work was supported by the State Grid Chongqing Electric Power Company & State Grid Chongqing Electric Power Company Tongliang Power Supply Branch, through the scientific project No.2020-2# & contract No. SGCQTL00SJJS2000193.

## References

1. Jiang Yuan, Li Qing, Feng Qian, Guo Lin, Cui Jiarui, Sha Guanglin. DC grid fault location and protection method based on BP neural network[J]. High Voltage Apparatus, 2020, 56(08): 23-28 .
2. Li Ke, He Jin, Cao Lucheng, Yang Fan, Zhou Shijin. Application of optimized BP neural network in fault location of distribution network[J]. Electrical Engineering Technology, 2019(19): 1-3.
3. Yang Kai, Yu Bo, Xiao Yanli, He Yongping, Wang Fengxiao. Distribution network project cost prediction based on GA-BP neural network [J]. Automation Instrumentation, 2019, 40(07): 91-93+99.
4. Sun Baohua, Chen Lei, Xia Dong, Han Tao. Design and application of intelligent operation and maintenance management and control platform for distribution network based on big data platform[J]. Electrical Automation, 2018, 40(06): 81-84.

5. Cheng Liang, Sun Qian. Risk early warning management and control of distribution network based on big data[J]. Shanxi Electric Power, 2020(03): 39-42.