Research on the trend forecasting model of power communication network operation

Shi Jian [1, a], Wu Hai-yang [2, b, *]

[1] State Grid Electric Power Research Institute, Nanjing, 210003, China

[2]Jiangsu Provincial Power Company, Nanjing, 210024, China

[a]email: 13675159059@139.com, [b]email: wuhaiyang@foxmail.com, [*]corresponding author

Keywords: Power transmission network; trend forecast; Performance characteristic

Abstract. This paper deeply studied the performance data of the power transmission network, selected indicators parameter in the characterization of the running state of the network. By analyzing the factors that affect the time series, a trend forecasting algorithm of network performance based on time series decomposition is constructed. It can make quantitative estimation for the future development trend of communication network. The model and algorithm of the paper are verified by the actual performance data. The results show that the forecast value is in good agreement with the actual monitoring data, and it has good practicability, which can provide technical support and judgment basis for the pre-warning of power communication network.

Introduction

The real-time diagnosis of power communication network operating state is directly related to the security and stable operation of power system. With the overall propulsion power enterprise smart grid and the "three five" system, more and more power grid operation control business and enterprise management information services required for transmission via power communication network. This makes the power system depend on the communication network continuously. Single point of failure of communication networks might adversely affect the security of the power system production.

Traditional power communication network monitoring and diagnosis is based mainly on the professional network management system prompts alarms. But for those signs is not obvious, there may be hidden dangers of performance information is the lack of effective diagnostic tool. With the power communication network scale and professional NMS continues to expand, this original real-time monitoring and post-processing of diagnostic method has been unable to meet the needs of power communication lean management.

Communication operation and maintenance personnel not only need to know the current running state of the communication network, but also want to know the running trend of the network and future running status change. So that maintenance personnel can adjust the operation mode of communication equipment reasonably, and extend the maintenance interval of the communication equipment scientifically, and ultimately can improve the efficiency of communication equipment [1]. Therefore, the operating status of power communication network prediction can be found potential failure of the communication equipment in advance. This has very important significance for reducing maintenance cost, improving the safety and reliability of communication network, and reducing the economic loss caused by network failure.

In this paper, we use the professional network management system to provide the performance parameters of the network operation state, construct the network trend forecast model, predict the running state of the communication network, and provide technical support for the maintenance mode based on the equipment state maintenance.

The Basic Theory and Method of Timing Analysis

Time series analysis by analyzing the correlation between different time variables, revealing the

structure and rules related to time series, so as to understand the inherent characteristics of the time series system, master data internal system and external relations, from the system's past values to predict and control the future value.

In this paper, the time sequence analysis method of optical transmission network is mainly used in two aspects. 1) Forecast analysis. According to the performance index, the statistical forecasting model is established, and the value of the performance index in the future is predicted. 2) Identification analysis. According to the records of two different time series, a statistical analysis model is established to judge whether it has the same property, or to judge whether it has certain properties according to the record value of a time series.

Set up a set of random variables $X_1,\ldots,X_n$, the sequence of which is called a random sequence, represented by $X_t(t=1,2...n)$.If the index is an integer variable, which represents the amount of increase of the time interval, such as the first time t, the first day t, this random sequence is called time series[2-4].

The time series of the characteristic values can be obtained by arranging the numerical value of a characteristic quantity. In reality, the time series of optical transmission network performance parameters are influenced by many factors. On the one hand due to accidental, non-decisive influence of random factors demonstrate its stochastic volatility and irregularities. On the other hand due to some fixed factors or cyclical factors, the decisive impact of the changes in the trend and a certain regularity. For example, the performance parameters of the optical transmission network are subject to some random disturbances in the process of monitoring, which makes the data of the characteristics of the optical transmission network show some random fluctuations. At the same time, in the process of monitoring, it is affected by some periodic operating conditions, and the value of it is also showing periodic fluctuation. When the optical transmission network is affected by a potential failure, the performance parameters of the optical transmission network will show a trend of some kind of continuous rise or fall.

The purpose of introducing the time series decomposition method is to analyze and distinguish the factors that affect the numerical value of the characteristics. By analyzing the regularity of the sequence variation, the long-term trend of the potential failure of communication equipment is revealed, and the future development trend is predicted.

According to the factors that affect the time series of the state monitoring of the communication network, the changes of time series can be decomposed into three types [5]. Trend changes: It refers to the phenomenon by fixed factors will change over time, exhibit a certain direction toward a steady rising, falling or steady trend. Cycle changes: It refers to the phenomenon of cyclical factors will be demonstrated by a fixed period of cyclical fluctuations. Similarly, periodic changes due to certain factors can lead to periodic fluctuations in the data. Random changes: It refers to the phenomenon affected by accidental factors exhibit irregular fluctuations.

The primary task of time series analysis is to decompose the collected performance parameters into a trend term, periodic term and random term. The three different terms of decomposition were established different regression model. Estimated by the known performance parameter data, and finally ultimately predict trends based on time series. What method to analyze and evaluate the change of each factor in time series is mainly determined by the relationship between the three factors. There are mainly two kinds of additive relation and multiplicative relationship in the relationship of time series. So the additive model or multiplicative model is formed. Because the factors affecting the performance parameters of the communication network monitoring are independent of each other. Therefore the forecast should be obtained the sum of the three factors. Additive model can be used in time series analysis.

$$X_t = M_t + S_t + I_t \tag{1}$$

Including: X_t is the original data item, M_t is the trend term, S_t is the periodic term. And satisfies $S_{t+d} = S_t$, $\sum_{j=1}^{d} X_j = 0$, I_t is random item.

Trend Forecasting Model Based On Time Series Analysis

A.trend term forecasting model

For trend term, multiple linear regression models can be used to estimate and predict.

The general form of multiple linear regression model are:

$$y = \mu_0 + \mu_1 x_1 + \cdots + \mu_m x_m + \varepsilon \quad (2)$$

Including: $\mu_0, \mu_1 \ldots\ldots \mu_m$ is an unknown parameter for m+1, called the regression coefficient. y is called the dependent variable. $x_1, x_2 \ldots\ldots x_m$ is a general variable that is acquired by N variables, called the explanatory variable.

When m = 1 is a linear regression model. when m$\geqslant$2 is a multivariate linear regression model. ε represents the random error.

For N group monitoring data obtained, group i can be expressed as $y_i = x_{i1},\ x_{i2} \ldots\ldots x_{im}$. Including i=1,2,3……n. So the linear regression model can be expressed as:

$$\begin{cases} y_1 = \mu_0 + \mu_1 x_{11} + \cdots + \mu_m x_{1m} + \varepsilon_1 \\ y_2 = \mu_0 + \mu_1 x_{21} + \cdots + \mu_m x_{2m} + \varepsilon_2 \\ \qquad\qquad \cdots\cdots \\ y_n = \mu_0 + \mu_1 x_{n1} + \cdots + \mu_m x_{nm} + \varepsilon_n \end{cases} \quad (3)$$

Due to the presence of residual between the actual value and the estimated value $e_i = y_i - (\mu_0 + \mu_1 x_{i1} + \cdots + \mu_m x_{im} + \varepsilon_i)$. Using the least square method, making the minimum residual sum of squares, can improve the accuracy of estimates:

$$\sum e_i^{\,2} = \sum (y_i - (\mu_0 + \mu_1 x_{i1} + \cdots + \mu_m x_{im} + \varepsilon_i))^2 \quad (4)$$

Its essential conditions are:

$$\frac{\partial \sum e_i^{\,2}}{\partial \mu_j} = 0,\ j = 0,1 \cdots m \quad (5)$$

After calculation, the equation (4) can be converted to:

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix}' \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_n \end{bmatrix} = X^{'} e = 0 \quad (6)$$

To estimate the value of $\hat{y} = x\hat{\mu}$, the estimated value and the actual value of the deviation is $y - \hat{y} = e$. So $y = \hat{y} + e = x\hat{\mu} + e$. While multiplying x' by the left and right, we can be obtained $x'y = x'x\hat{\mu} + x'e$. And $x'e = 0$. therefore, $\hat{\mu} = (x'x)^{-1}x'y$.

The least square linear regression estimation of the $\hat{\mu}$, the necessary inspection and evaluation is carried out. These include: the fit of the regression equation determined the estimated standard error of regression equation hypothesis testing, hypothesis testing and multiple linear discriminant , etc.. Calculate the estimated value $\hat{\mu}$ based on samples collected performance parameters. Verified by actual values, the final sample regression equation is obtained.

$$\hat{y} = \hat{\mu}_0 + \hat{\mu}_1 X_1 + \hat{\mu}_2 X_2 + \cdots + \hat{\mu}_m X_m \quad (7)$$

Since $\hat{y}$ is an unbiased estimate of $E(y)$, $\hat{y}$ can be used as the final forecast.

A. periodic term forecasting model

Periodic term has a periodic variation of characteristics, and in a certain period of time, there is a constant repetition of the regular. Therefore, the point on the same phase will change up and down in the vicinity of a particular value. The periodic term S_t is decomposed, and the data points on the same phase constitute the sequence.

Set sequence of the periodic term is $\{T(t), t = 1, 2, \cdots N\}$[6], Cycle is $T = h$, The same phase of the data points in the same column. Then the original sequence can be transformed into the following matrix:

$$\begin{bmatrix} T(1+0\times T) & T(2+0\times T) & \cdots & T(h+0\times T) \\ T(1+1\times T) & T(2+1\times T) & \cdots & T(h+1\times T) \\ \vdots & \vdots & \ddots & \vdots \\ T(1+(\frac{N}{T}-1)\times T) & T(2+(\frac{N}{T}-1)\times T) & \cdots & T(h+(\frac{N}{T}-1)\times T) \end{bmatrix} \quad (8)$$

To split the matrix (8) into h different phases of the sub sequences by a column vector, and then complete the reconstruction of the period. Modeling each sub sequence, by simulating the variation of the same phase in different periods, the data values in the next cycle can be predicted, and the multi-step prediction is transformed into a single step prediction.

B. Random-term prediction model

In addition to the trend term and periodic term in time series, there may be some kind of equilibrium volatility, which is usually stationary sequence. For stationary series, the time series analysis method based on the Auto-Regressive and Moving Average Model (ARMA Model) can be used to model and predict . ARMA as a limited parameter model, just make sure a finite number of parameters, we can completely determine the model [7-10]. ARMA model is used to calculate the predictive value of a random term based on the t of the p moments before the q values of the historical value and the linear calculation of the error values are obtained. ARMA model is used to calculate the predictive value of a random term, which is obtained by the linear calculation based on the he p historical values and q error values of a t time. ARMA (p, g) model can be described as:

$$V(t+1) = \sum_{i=1}^{p} \varphi_i V(t+1-i) + \sum_{j=1}^{q} \theta_j \varepsilon(t+1-j) \quad (9)$$

Formula (9) $V(t+1)$ is the next time the predictive value of a random item. φ and θ are not equal to zero the undetermined coefficient. p and q are respectively ARMA model auto regressive order and moving average order. $V(t+1-i)$ is a random item measurement value at t time. $\varepsilon(t+1-j)$ is the prediction error term random items before time t. As a result, the most important parameter in the model is to determine the value of p and order q.

For the order q, when the time series of the autocorrelation coefficient in a certain d order after nearly 95% of the values are in the range of 2 times the standard deviation, and the self-attenuation coefficient of fluctuation processes in small value, the model order can be considered value for this d. For the value of the model parameter p, the least square regression method is used to estimate the N data. The residual value is obtained

$$e = V(t+1) - \sum_{i=1}^{p} \varphi_i V(t+1-i) + \sum_{j=1}^{q} \theta_j \varepsilon(t+1-j) \quad (10)$$

Among them, p=1,2,... , N, $V(t+1)$ is the actual value for t+1 time. The residual e is the deviation of the actual value and the predicted value. N is the number of samples. According to the formula (4) solution with Trend Prediction least squares regression method can be obtained for $\hat{\beta} = (x'x)^{-1}x'y$.

After the model parameters p and the value of the order q are determined, it can be predicted according to the historical data.

Algorithm Implementation and Analysis

A. Algorithm implementation

The prediction step of the trend prediction algorithm based on time series are mainly the following steps:

- Network performance parameters selection. According to the relevant authority national standards and industry standards, the field of experimental data, a large number of literature data and expert advice, choose to reflect the network operation state of network management performance parameters.
- Network management performance data acquisition. To obtain the specified network performance data from the north to the interface of the optical transmission equipment network management.
- Time series. In the sampling period, the performance data of the network management performance data is collected, and the time series of the characteristic values of the parameters are arranged according to certain time interval.
- Time series decomposition. By analyzing the sample of time series, the trend term, periodic term and random term of time series are decomposed.
- Calculation of partial forecast value. For three different types of decomposition, the forecast values are estimated respectively according to their respective forecasting models.
- Final forecast value calculation. According to the time series addition model, calculate the final forecast value, and cross validation with the real value.

B. Case analysis

In optical transmission equipment, there are a lot of performance parameters which affect the operation quality of communication network and the service quality. Including background block error, error code block, bit error rate, unavailable seconds, FEC, FEC, CRC, pointer modification count, protection switching count, temperature, input optical power, optical output power and laser bias current rate, as in figure. Among them, bit error rate and optical power is an important index to measure the quality of signal transmission in communication network.

The optical power transmission value of a practical power communication net optical transmission device is used as the sample. Within a year, 365 sets of data were collected for 24 hours.

Based on the time series of the trend prediction algorithm steps are as follows:

The trend term M_t is decomposed by the sliding average algorithm, and the prediction data of the trend term is obtained by using the least square regression method. The order is the 8 time.

Periodic term S_t is obtained by means of multi period superposition, and then the prediction of subsequent periodic data is carried out using continuation method.

Random terms I_t was obtained by $X_t - M_t + S_t$, and then using ARMA (p, g) model to obtain the random data.

Through time series addition model $X_t(l) = M_t(l) + S_t(l) + I_t(l)$, the final forecast value was obtained. The accuracy of the forecasting method is verified by cross correlation.

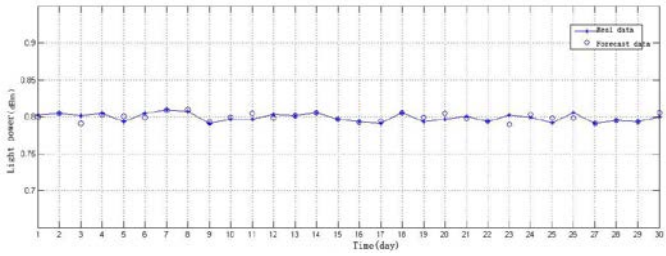

Fig. 1. Forecast data with actual data comparison chart

As shown in Figure 2, through cross correlation, the experimental results show that the trend of prediction and the actual monitoring trend are in good agreement, and have good practicability.

Conclusion

The existing fault diagnosis idea is based on the fault symptom, using the characteristic parameters of the condition monitoring and fault diagnosis of the fault knowledge base to classify and identify the equipment failure, to judge the location of the equipment, the nature and the extent of the fault, and the development trend of the equipment failure.

In this paper, we start with the characteristics of the normal operating state of the communication network, and establish the communication network energy trend forecast model based on the time series. Real time performance data of communication network is used to verify the model and algorithm. The results show that the forecast trend has a good agreement with the actual monitoring trend, which can meet the potential anomaly of the early warning network, and has good practicability and application prospect.

References

[1] Zhu Li, QinLei, Xue Kouying. A Novel BP Neural Network Model for Traffic Prediction of Next Generation Network[A]. International Conference on Natural Computation[C], Tianjin China IEEE, 2009: 3 2−3 8.

[2] LI Shu-hui, WUNSCH D C,O'HAIR E A. Using neural networks to estimate wind turbine power generation[J]. IEEE Transactions on Energy Conversion, 2001, (03): 276-282

[3] Billinton R, CHEN H, GHAJAR R.. Time-series models for reliability evaluation of power systems including wind energy[J]. Microelectronics Reliability, 1996, (09): 1253-1261.

[4] Wang P, BILLINTON R. Time-sequential simulation technique for rural distribution system reliability cost/worth evaluation including wind generation as alternative supply[J]. IEE Proceedings-Generation Transmission and Distribution, 2001, (4): 355-360.

[5] Lv Lintao, Wang Peng, Li Junhuai. Research on the Trend Analysis and Predictive Algorithm Based on Time Series[J]. Computer Engineering and Applications. 2004. 19: 173.

[6] Song Xianlei, Liu Yezheng, Chen Sifeng. Seasonal Time Series Forecasting Based on Seasonality Method Selection[J]. Computer Engineering, 2011, 37(21): 132. 135.

[7] Ding Ming, Zhang Lijun, Wu Yichun. Wind speed forecast model for wind farms based on time series analysis[J]. Electric Power Automation Equipment, 2005, 25(8): 32−34.

[8] Box G E P, Jenkins G M. Time series analysis: Forecasting and control [M]. San Francisco: Holden-Day, 1976.

[9] Gao Junfang, Wu Qing, ARMA's modeling the time series and its application[J]. Journal of Shanghai University of Engineering Science. 1996,(10):68-73.

[10]Xu Guohui, Yu Chunli. A study on time series analysis method[J]. Journal of Guangzhou University(Natural Science Edition) . 2003,(06):556-559.