Research on Short-Term Load Forecasting Based on PCA-GM

Hai-Hong Bian[1,2(✉)], Qian Wang[1], and Linlin Tian[3]

[1] Nanjing Institute of Technology, Nanjing 211167, Jingsu, China
llq20180l@163.com
[2] Jiangsu Collaborative Innovation Center for Smart Distribution Network, Nanjing 211167, Jingsu, China
[3] School of Information and Control, Shenyang Institute of Technology, Shenyang 113122, China

Abstract. In this paper, a short-term load forecasting model based on PCA dimensionality reduction technology and grey theory is proposed. After the correlation analysis between meteorological factors and load indicators, the data is carried out by combining PCA dimensionality reduction technology and grey theoretical load forecasting model. In this paper, the validity of the load data verification model in a western region is selected. The analysis of the example shows that compared with the general gray prediction model GM (1, 1), the accuracy of the model prediction result is much higher, which proves the model. Effectiveness and practicality.

Keywords: Short-term · Load forecasting · PCA-GM

1 Introduction

Under the current situation of increasingly tight international power supply situation, accurately predicting the power load and rationally arranging the power system production and power supply plan are of great significance to the safe and reliable operation of the power system, the sustained development of the national economy and the normal life of the residents [1]. There are many methods for power load forecasting, such as gray forecasting, regression analysis and forecasting, traditional trend analysis and exponential smoothing. These algorithms have their own advantages and disadvantages: (1) From the applicable conditions, regression analysis and trend analysis are devoted to the research and description of statistical laws, and are suitable for large-sample and consistent predictions of past, present and future development models; exponential smoothing The method is to use the principle of inertia to extrapolate the growth trend, while the gray model method seeks the law by sorting the original data, which is suitable for analysis and prediction under the condition of poor information; (2) from the data form adopted, gray The theoretical model uses the generated numerical sequence modeling. The regression analysis method and the trend extrapolation method are all based on the original data modeling. The exponential smoothing method directly predicts the future value by exponentially weighting the original data; (3) from the applicable time In terms of classification, regression analysis and trend

Y.-D. Zhang et al. (Eds.): ICMTEL 2020, LNICST 327, pp. 169–177, 2020.
https://doi.org/10.1007/978-3-030-51103-6_15

extrapolation are more suitable for medium and long-term predictions. For short-term predictions, it is more suitable for grey theory models. Although the traditional grey theoretical model is more efficient and simpler than other predictive models, it does not predominate when using long data series predictions. The data columns are too long, the system is subject to many disturbances, and the instability factors increase. On the contrary, the accuracy of the model is reduced, and the credibility of the prediction result is also reduced.

Because of the above problems, this paper improves the traditional grey theory algorithm, and considers the complexity and multi-dimensionality of meteorological factors that affect the load. A short-term load forecasting model based on PCA dimensionality reduction and grey theory is proposed. A correlation analysis and PCA dimension reduction method are used to analyze the degree of correlation between meteorological factors and load indicators, and the factors with low correlation degree are eliminated, and the multidimensional variables are reduced to one dimension by PCA dimensionality reduction technology, simplify the prediction model and reduce the amount of calculation.

2 Short-Term Load Forecasting Under Grey Theory

2.1 GM (1, 1) Model

The grey system theory is based on the concept of associative space, smooth discrete function and other concepts to define gray derivatives and gray differential equations [2]. By processing the original data, the system changes the law, and then the discrete data series is used to build the dynamic model of the differential equation. This is the intrinsic gray [3]. The basic model of the system, and the model is approximate, non-unique, so this model is called the gray model, denoted as GM (Grey Model).

The gray model in the general sense is GM (u, h), which means that u-order differential equations are established for h variables. The model used for prediction is generally GM (u, 1), and the most practical application is GM (1, 1), so only the GM (1, 1) model needs to be established.

The above standard "(0)" indicates the original sequence [4], and the superscript "(1)" indicates the cumulative generation sequence. The main steps of modeling and predicting GM (1, 1) are as follows:

(1) Let the original sequence be

$$X^{(0)} = \left[x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \cdots x^{(0)}(n)\right] \tag{1}$$

(2) Accumulating the original sequence (AGO) generation

$$x^{(1)}(j) = \sum_{i=1}^{j} x^{(0)}(i), j = 1, 2 \cdots n \tag{2}$$

Can get

$$x^{(1)}(i) - x^{(1)}(i-1) = x^{(0)}(i), i = 1, 2, 3 \cdots n \tag{3}$$

(3) Establish the corresponding differential equation as

$$\frac{dx^{(1)}}{dt} + mx^{(1)} = n \tag{4}$$

Where, m is the development coefficient and n is the ash dosage [5]. The effective interval of m is that the matrix composed of m and n is the grey parameter $\hat{m} = (m \quad n)$. only m and n are required, then $x^{(1)}$ can be obtained, and then the prediction value of $x^{(0)}$ can be obtained.

(4) Build the mean of the accumulated generated data B and the constant term vector Y_n

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}, Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{5}$$

(5) Solving the ash parameter $\hat{m}$ by the least squares method, then

$$\hat{m} = (B^T B)^{-1} B^T Y_n$$

(6) Substituting the ash parameter into the formula (1.4) and solving it,

$$\hat{x}^{(1)}(i) = [x^{(0)}(1) - \frac{n}{m}]e^{-m(i-1)} + \frac{n}{m}, i \geq 1 \tag{6}$$

(7) The above results are reduced and reduced to obtain the predicted value

$$\hat{x}^{(0)}(i) = \hat{x}^{(1)}(i) - \hat{x}^{(1)}(i-1), i \geq 2 \tag{7}$$

The establishment of the GM (1, 1) model generally requires a series of tests. Once the established GM (1, 1) model fails the test by the three methods of residual, correlation and posterior difference, the model must be tested. Corrected. Compared with the traditional methods of statistical analysis through a large number of samples, the superiority of gray system modeling is reflected in two aspects [6]. One is that the former requires a large amount of raw data, and its accuracy can be guaranteed, while the latter does not have such demanding requirements [7]. The second is that the latter generally uses a certain way to generate and process the original data (such as accumulation generation and subtraction generation), and organizes the disordered raw data into regular generation data, thereby weakening the randomness of the original random

sequence. A smooth discrete function is obtained, which is further modeled based on these generated data. Therefore, the gray model has the characteristics of less information required for modeling and higher modeling accuracy.

2.2 Analysis of the Influence Degree of Meteorological Factors on Load

There are many factors that affect short-term load forecasting. However, we are unable to assess the extent to which these factors affect the load. Differences in load structure and climate will also change the form of the load. Therefore, short-term load is difficult to predict to a certain extent, which is the main solution to short-term load forecasting. The results show that meteorological factors are the most important factors affecting the load. Temperature is considered to be one of the main influencing factors [8]. In addition to the highest temperature of the day, the effects of average temperature and minimum temperature on the load should also be considered.

The influence of meteorological factors on power load forecasting is very complicated, and there are often interactions between different meteorological factors. This requires further analysis of the coupling strength between multiple meteorological factors and load forecasting [9]. The larger the correlation coefficient, the higher the correlation between the two; the lower the correlation coefficient, the lower the correlation.

The formula for calculating the correlation coefficient is as follows:

$$R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{8}$$

In the above formula: R is the correlation coefficient; $\bar{x}$ and $\bar{y}$ are the average values of meteorological indicators and load indicators [10], respectively. Among them, the larger R is, the higher the correlation degree is. When $|R| > 0.8$, x is highly correlated with y; when $|R| < 0.8$, x is associated with y low; when $|R| = 1$, x and y are completely related.

The correlation data between the highest temperature, average temperature, minimum temperature, and rainfall data of a certain week in July 2018 in the western part of July 2018 was calculated and correlated with the load data of the day. Four correlation coefficients were obtained, which were denoted by $R_1 \sim R_4$, such as Table 1 shows.

Table 1. Correlation coefficient result table

R_1	R_2	R_3	R_4
77.4%	81.0%	39.8%	11.0%

From the correlation coefficient, the correlation between rainfall and load data is small; due to the small amount of rainfall in summer, the relationship between rainfall

and load in summer is also very small. Therefore, when constructing short-term load forecasting using the grey theory method, it is only necessary to consider the influence of the highest temperature, average temperature and minimum temperature on the load data.

3 Improved PCM-Based GM Model

PCA (Principal Component Analysis) is a common dimensionality reduction method. It focuses on mitigating dimensional disasters, minimizing information loss while compressing data, and understanding simple multidimensional data. When understanding feature extraction and processing data, the problems involving high-dimensional feature vectors tend to fall into dimensional disasters. As the data set dimension increases, the number of samples required for algorithm learning increases exponentially. When processing a large number of highest temperature, daily average temperature, minimum temperature and load data after correlation analysis [11], the data is sparsely high due to the higher dimension, and the same data set is explored in the high dimensional vector space. It is more difficult to explore sparse data sets.

Principal component analysis, also known as the Karhunin-Lough transform, is a technique for exploring high-dimensional data. PCA is commonly used for the exploration and visualization of high-dimensional data sets, and can also be used for data compression, data pre-processing, and so on. PCA can synthesize highly-dimensional variables that may be correlated into linearly independent low-dimensional variables, which are called principal components. The new low-dimensional datasets retain the variables of the original data as much as possible, and can effectively extract the most valuable data. Save a lot of time to achieve uniformity and coordination of load forecasting results.

There are two common methods for achieving PCA dimensionality reduction. One is based on the eigenvalue decomposition covariance matrix, and the other is based on the SVD decomposition covariance matrix. This article selects the latter, the specific steps are:

(1) Input: data set $X = \{x_1, x_2, x_3 \cdots x_n\}$, need to be reduced to k-dimensional;
(2) The average value is the average of each feature minus its respective average value;
(3) Calculate the covariance matrix $\frac{1}{n}XX^T$;
(4) Calculating the eigenvalues and eigenvectors of the covariance matrix $\frac{1}{n}XX^T$ by SVD;
(5) Sorting the eigenvalues from small to large, selecting the largest k, and then composing the corresponding k eigenvectors as the row vector to form the eigenvector matrix P;
(6) Convert the data into a new space constructed by k feature vectors, which is $Y = PX$.

In the PCA dimensionality reduction, the largest k eigenvectors of the sample covariance matrix need to be found, so that the matrix composed of the largest k eigenvectors can be used for low dimensional projection dimensionality reduction. In the process of using MATLAB to achieve PCA dimensionality reduction, a combination of SPE (square prediction estimation, also known as Q statistic) and T2 (principal component analysis) is used to monitor the fault. The Q statistic indicates the degree of deviation of the test value from the principal at this time. If the Q statistic is too large, it indicates that an abnormal situation has occurred in the process, so it is usually used to display the abnormal situation; the T2 statistic indicates the principal component analysis. Statistics. It can be seen from Fig. 1 and Fig. 2 that the Q statistic and the T2 statistic are basically within the threshold, indicating that the dimension reduction process is operating normally.

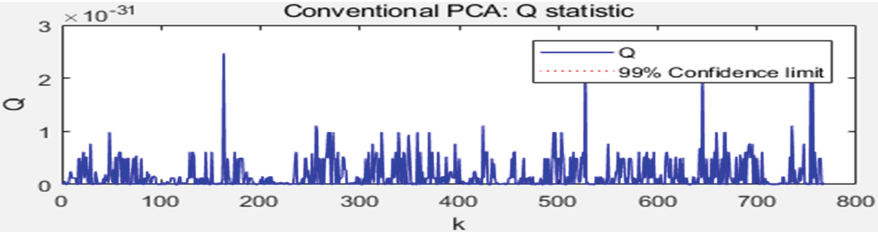

Fig. 1. Q Statistics

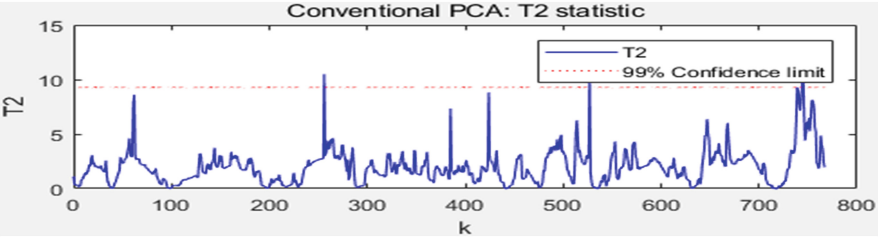

Fig. 2. T2 Statistics

4 Experiment and Result Analysis

In order to verify the accuracy of the load forecasting model based on PCA dimensionality reduction and grey theory proposed in this paper, the load data from July 27 to August 2, 2018 in the western region was taken as a sample, and in August 2018. On the 3rd day, as the forecast date, the 96 point load of the day was predicted. The specific process is shown in Fig. 3:

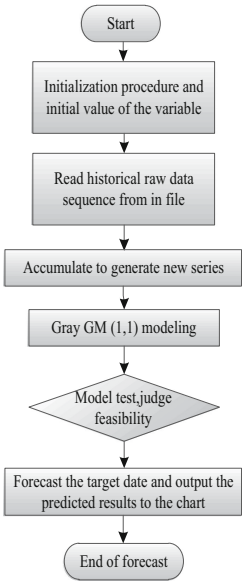

Fig. 3. Flow chart based on PCA and grey theory load forecasting

In this example, after the analysis of the correlation degree of meteorological factors, the influence of rainfall on the load index is removed, and only the three factors affecting the daily maximum temperature, the daily average temperature and the daily minimum temperature are considered; The maximum temperature, the daily average temperature and the daily minimum temperature are used for dimensionality reduction. Finally, the gray system theory is used to predict the power system load value. The simulation results are shown in Fig. 4 and Fig. 5.

Figure 4 shows the comparison between the predicted 96 points and the actual 96 points. The red line is the predicted value and the blue line is the actual value.

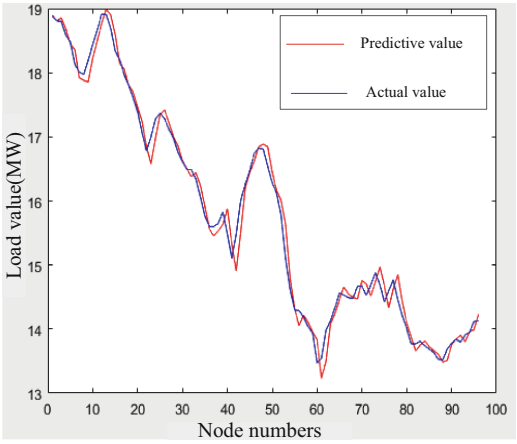

Fig. 4. Comparison of predicted and actual values (Color figure online)

It can be seen from the analysis of the above figure that the change range of the actual value is 18.8 MW–14.2 MW, and the change range of the predicted value of the method in this paper is 18.8 MW–14.3 MW. It can be seen from the observation that the predicted value and the actual value of the research method in this paper have a high degree of fit, which shows that the prediction accuracy of the method in this paper is high, and it can realize the accurate prediction of the short-term load. The reason why this method has high prediction accuracy is that the improved PCA GM based prediction model considers the complexity and multi-dimensional of climate factors, improves the accuracy of load forecasting, overcomes the shortcomings of traditional prediction methods, so it has high prediction accuracy.

Figure 5 clearly shows the error between the predicted value and the actual value.

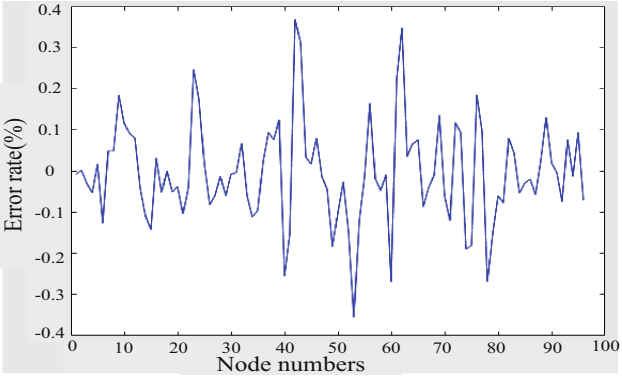

Fig. 5. Error between predicted and actual values

It can be seen from the analysis of Fig. 5 that the prediction error of this method varies from −0.36% to 0.37%, the prediction error is small, and the change amplitude of the prediction error is small, which further proves that this method can realize the accurate prediction of short-term load, and verifies the comparison results of Fig. 4. The reason is that by calculating the correlation between meteorological factors and load index, the influencing factors with high correlation are selected. On this basis, the multi-dimensional meteorological factors are reduced to one dimension, and the GM (1, 1) grey theoretical model is used for prediction, which effectively solves the problems existing in the traditional methods, reduces the prediction error and improves the prediction accuracy.

5 Conclusion

(1) By calculating the degree of correlation between meteorological factors and load indicators, the influencing factors with high correlation degree are selected. On this basis, the multi-dimensional meteorological factors are reduced to one-dimensional, and the GM (1, 1) gray theoretical model is used for prediction.

The error between the predicted value and the actual value obtained after bringing in the sample is only within ±0.04%, which is closer to the actual situation, which proves that the model is reasonable.

(2) Considering the complexity and multidimensionality of climatic factors, the improved PCA-GM-based prediction model improves the accuracy of load forecasting and overcomes the shortcomings of traditional forecasting methods.

(3) Compared with the traditional GM (1, 1), the result is more accurate, and the calculation model is simplified, the calculation amount is reduced, the prediction accuracy is improved, and the validity of the model is proved.

In the load forecasting process, only four influencing factors of maximum temperature, average temperature, minimum temperature and precipitation are considered. In the actual situation, there are still many influencing factors to be explored. Uncertainty and controllability in load forecasting The variables are more complicated and there are still many problems to be further studied.

Fund Projects. The project was supported by the Open Research Fund of Jiangsu Collaborative Innovation Center for Smart Distribution Network, Nanjing Institute of Technology (No. XTCX 201807).

2019 Jiangsu Province Graduate Practice Innovation Plan (SJCX19_0519).

References

1. Gao, C., Li, Q., Su, W., et al.: Temperature correction model considering the effect of accumulated temperature in short-term load forecasting. Trans. China Electrotech. Soc. **30**(4), 242–248 (2015)
2. Liu, B., Ma, J., Li, X.: A topic representation model of "Feature Dimensionality Reduction" text complex network. Data Analysis Knowl. Disc. **1**(11), 53–61 (2017)
3. Li, Q.: New characteristics of load characteristics analysis and load forecasting in smart grid (2014)
4. Zhang, Q.: Short-term load forecasting considering temperature accumulation effect (2014)
5. Wang, W., Wang, B., Yu, H., et al.: Maximum load forecasting of power grid based on load decomposition and grey theory. Zhejiang Electric Power **26**(8), 48–53 (2018)
6. Cheng, D., Liu, J., Guo, W., et al.: Power load forecasting of central china power grid based on accumulated temperature effect. Meteorol. Sci. Technol. **46**(269(04)), 186–193 (2018)
7. Zhang, H., Li, W., Xiang, C., et al.: Short-term load forecasting based on temperature cumulative effect and grey correlation degree. Electric Autom. **26**(3), 12–19 (2019)
8. Cheng, Z.: Research on short-term load combination forecasting method considering accumulated temperature effect (2010)
9. Zhang, Q., Wang, Y., Lu, Y.: Study on summer daily maximum load forecasting considering accumulated temperature effect. Power Demand Side Manag. **26**(6), 1–5 (2013)
10. Renyuan, Z.: Analysis of the cumulative effect of temperature in Shanghai electricity load forecasting. Modern Electric Power **35**(2), 38–42 (2018)
11. He, W., Ye, P., He, N., et al.: Review of power system saturation load prediction research. J. Shenyang Inst. Technol. (Natural Sci. Ed.) **13**(4), 340–346 (2017)