

Online Compression Reconfiguration-Based Load Forecasting Method for Distribution Grid Power System



Wenqi Huang, Lingyu Liang, Shang Cao, Xiangyu Zhao, Huanming Zhang,
and Hanju Li

Abstract The current conventional power system load forecasting method mainly outputs forecasting results by constructing a time model, which leads to poor forecasting results due to the lack of effective extraction of feature data of load signals. In this regard, an online compression and reconstruction-based load forecasting method for distribution network power systems is proposed. By introducing the concept of particle swarm ensemble, the discrete situation of power load signal data particles is characterized, and data normalization is carried out, and the load signal data is compressed and reconstructed. The maximum information coefficient is calculated and the load data features are extracted by combining the influencing factors, and finally a hybrid prediction model is constructed and the model is solved. In the experiments, the designed method is verified for its prediction effect. The experimental results show that the designed method has a good fit between the prediction results and the actual load curve, and has a good prediction performance.

Keywords Online compressive reconstruction · Power systems · Load forecasting · Sparse dictionaries

1 Introduction

Distribution network power system load forecasting is mainly through the analysis of historical operation data of the distribution network, combined with forecasting algorithms, to achieve the prediction of the expected load situation of the distribution system. By predicting the expected power load of the distribution system, the operating conditions of the power system can be grasped, thus helping the staff to make correct maintenance decisions and realize the effective allocation of power resources [1]. If the upcoming power load can be accurately predicted before the

W. Huang · L. Liang · S. Cao (✉) · X. Zhao · H. Zhang · H. Li
China Southern Power Grid, Digital Grid Research Institute, Guangzhou 510663, China
e-mail: caoshang@csg.cn

power peak period, the power system can be pre-adjusted to meet the power demand at different time periods and ensure the smooth operation of the power system. Distribution system load forecasting can be divided according to the forecasting cycle, with different forecasting objectives and techniques corresponding to different forecasting cycles. It is also susceptible to the influence of variable factors such as weather, which leads to the instability of the prediction results. At the same time, most of the current power systems have a large volume of operational data, and the analysis of historical data is a large amount of engineering work, so the use of conventional methods for load forecasting not only easily leads to deviations in the forecasting results, but also easily affects the forecasting efficiency, which cannot meet the short-term forecasting needs. In this paper, we propose an online compression and reconstruction-based load forecasting method for distribution network power system, which aims to improve the forecasting efficiency and reduce the transmission of edge data by compressing and reconstructing the operational data, so as to improve the operational efficiency of the forecasting algorithm.

2 Online Compression Reconstruction-Based Load Signal Data Preprocessing for Distribution Network Power Systems

The specific expression of the search space is shown as formula (1) [2–4].

$$D(t) = \frac{1}{mL^2} \sum_{i=1}^m \sqrt{\sum_{d=1}^n (p_{id} - p_d)^2} \quad (1)$$

where m represents the number of power system load signal particles in the search space, L^2 represents the scale of the search space, n represents the three-dimensional vector constant of the search space, p_{id} represents the node coordinate data of the i -th power system load signal particle in the d -th search space, and p_d represents the average of all node coordinate data[5–7]. The above expression of the search space $D(t)$ can characterize the dispersion degree of the power system load signal particles in a certain range. When the dispersion degree of the load signal particles in a certain range is larger, the value of $D(t)$ will be larger; when the load signal particles in a certain range show aggregated distribution, the value of $D(t)$ will be smaller. Therefore, the above formula can determine the distribution of load signal particles in the search space at a certain time, and the optimal solution search can be realized by selecting the most suitable search space, so as to ensure the uniform distribution of load signal particles in the search space of power system. After finishing the above adjustment, all the data in the collection space of load signal particles of power system under uniform distribution are normalized in this paper, and the data fluctuation range is adjusted, and the specific calculation formula is shown formula (2).

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

Among them, x_{\max} and x_{\min} represent the maximum and minimum values of the power system load signal particle population under uniform distribution, x represents the original load signal particle data, and x^* represents the particle signal data after processing. After the preprocessing of the power system load signal particle data, the normalized data is obtained, and the particle signal data is compressed online according to the characteristics of the load signal data in the sparse dictionary. In this regard, it is necessary to first select a suitable signal measurement matrix, which is mainly used to measure the signal scale, and also needs to present a mutually independent state with the sparse dictionary. In this paper, the Gaussian random matrix $\Phi_{m \times n}$ is chosen to measure the particle signal scale, which is calculated as formula (3).

$$y_t = \Phi x_t = \Phi D A s = \Theta s \quad (3)$$

where y_t represents the observed signal, D represents the sparse dictionary, A represents the perceptual matrix, and s represents the sparse signal obtained after the mapping process is completed. By using the above formula, the input power system load signal at time t can be projected, thus changing the signal dimension and obtaining the particle signal scale. For the sparse signal s , a constraint expression needs to be constructed to constrain it, and the specific calculation formula is shown formula (4).

$$\min \|s\| \text{ s.t. } y_t = \Theta s \quad (4)$$

The sparse signal s can be constrained by the above steps, thus converting the constraint problem into a dimensional transformation problem, and then the value of s is calculated by using the regularized matching algorithm, first inputting the training data into the K -SVD sparse dictionary for learning, and then outputting the sparse matrix F . The value of the sparse signal is obtained by mapping it one-by-one with the actual data, thus outputting the mapping result, and then the load signal is online compression and reconstruction, the specific implementation process is shown in Fig. 1.

3 Power System Load Signal Data Feature Extraction

Compared with the mutual information coefficient, the maximum information coefficient can effectively measure the linear relationship between variables and mine the linear relationship between data under different attributes and has a better ability to deal with discrete data. The higher the value of the maximum information coefficient, the higher the similarity of the characteristics between the data. In this regard,

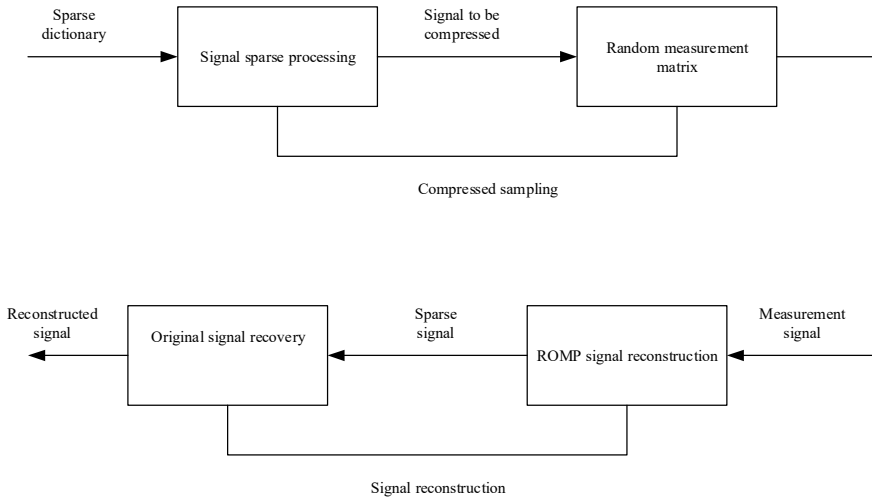


Fig. 1 Online compression and reconstruction process of power system load signal

this paper constructs the maximum information coefficient on the basis of mutual information coefficient, and the specific derivation formula is shown formula (5).

$$I(x, y) = \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{5}$$

where $p(x, y)$ represents the probability of joint distribution between load signal data x and load signal data y , $I(x, y)$ represents the mutual information coefficient, and $p(x)$ and $p(y)$ represent the probability of separate distribution of load signal data, respectively. After finding the mutual information coefficient, the value of the maximum information coefficient can be solved according to the construction of the sampling sample function, and the specific formula is shown formula (6).

$$\text{MIC}(x, y) = \max \frac{B(n)I(x, y)\sigma}{\log \min(x, y)} \tag{6}$$

where $B(n)$ represents the load signal sampling sample function and n represents the number of samples. σ represents the correlation coefficient, which characterizes the degree of correlation between two variables and corresponds as given in Table 1.

Considering that the temperature factor in the climatic conditions has a large influence on the load situation of the power system, seven characteristic factors including UV intensity, temperature, and humidity are selected in this paper, and the corresponding maximum information coefficients are calculated to compare the correlation degree between different climatic factors and the load situation of the power system, so that the characteristic factor with the largest correlation degree

Table 1 Correspondence of correlation coefficients

Correlation coefficient	Degree of correlation
0.00–0.19	Very low correlation
0.20–0.39	Low correlation
0.40–0.59	Moderate correlation
0.60–0.79	Highly correlated
0.80–0.99	Extremely high correlation

can be selected. The corresponding characteristic correlation coefficients of specific climatic factors are given in Table 2.

From the Table 2, it can be seen that the characteristic correlation coefficients of the two factors, UV intensity and whole hour, are large, which can be judged to be highly correlated with the load situation of the power system and can be used as the characteristic input variables of the load forecasting model. After the above analysis of the factors influencing the load situation, the maximum information coefficient values of different factors are set as the threshold values, and if the maximum information coefficient value of the factor is higher than the threshold value, it can be seen that the factor has a greater degree of influence on the load situation and can be used as the input value of the model, as formula (7).

$$\varpi = 1/N \sum_{i=1}^N MIC_i(Y) \tag{7}$$

where N represents the number of feature variables of load influencing factors, Y represents the power system load signal data sequence, and MIC_i represents the maximum information coefficient value under the i -th feature factor. The extracted features can not only be used as input data for the load prediction model of the power system, but also can be used to judge the degree of influence of the corresponding correlation coefficients, which can help to adjust the load operation of the power system.

Table 2 Correlation coefficients of characteristics corresponding to climatic factors

Characteristic	Description	Significance test	Maximum information coefficient value
UV	UV intensity	Significant	0.65
Tem	Temperature	Inconspicuous	0.38
Hum	Humidity	Inconspicuous	0.33
Winds	Wind speed	Inconspicuous	0.36
Cloud	Cloud cover	Inconspicuous	0.12
Dew	Dew point temperature	Inconspicuous	0.24
Hour type	Hour	Significant	0.78

4 Power System Load Forecasting Model Construction

Firstly, assume that the active values of neurons representing positive propagation in the BIGRU network structure at moment t and the active values of neurons representing negative propagation, thus obtaining the specific expression of the BIGRU network model as formula (8).

$$\begin{cases} c^{+t} = G(x^t, c^{+t-1}) \\ c^{-t} = G(x^t, c^{-t-1}) \\ h^t = w^{+t}c^{+t} + w^{-t}c^{-t} + \frac{b^t}{H} \end{cases} \quad (8)$$

where c^{+t} represents the implied values of neurons for positive propagation at time t , c^{-t} represents the implied values of neurons for negative propagation, x^t represents the input load characteristics data of the model, h^t represents the pooling information at time t , G represents the model output values, w^{+t} and c^{+t} represent the implied weights for positive and negative propagation, b^t represents the implied bias parameters, and H represents the modal confounding function.

5 Experiment and Analysis

5.1 Experimental Preparation

In order to prove that the online compression reconstruction-based distribution grid power system load forecasting method proposed in this paper is better than the conventional distribution grid power system load forecasting method in terms of actual forecasting effect, after the theoretical part of the design is completed, an experimental session is constructed to test the actual forecasting effect of the method in this paper. In order to ensure the experimental effect, two conventional distribution system load forecasting methods are selected for comparison, namely the data mining-based distribution system load forecasting method and the ELM-based distribution system load forecasting method. The specific experimental environment configuration is given in Table 3.

In this experiment, the historical data of power system operation under a large distribution network is retrieved as the dataset for the experiment, and the dataset is divided into two parts, which are used for algorithm training and experimental testing. The data sampling interval is set to 30 min, and a total of 64 load signal data sequences are constructed. The training load results of the model and the test load results are used as the standard to finally output the actual predicted load values. In order to improve the reliability of the experimental results, the load of the power system at different times of the day is selected as the test standard, and three methods

Table 3 Configuration of experimental parameters

Experimental environment parameters	Specific configuration
Simulation platform	MATLAB 2021a
Experimental environment	Intel Core(i) i5-9400f
Software system	Windows 7
Sampling interval	30 min
Sampling time points	4200
Training set data volume	3960
Test set data volume	240

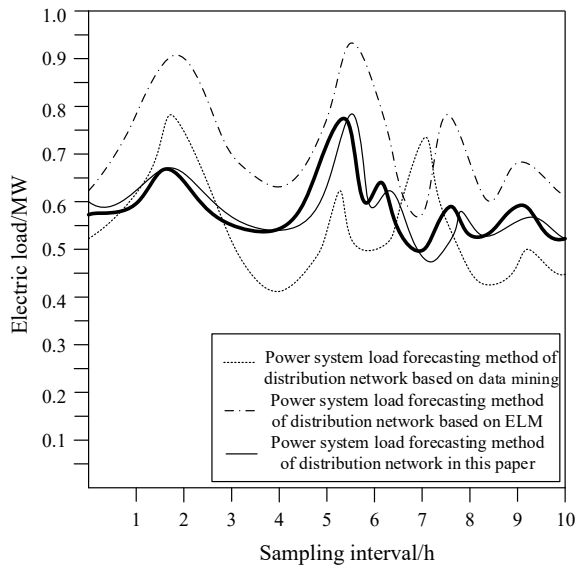
are used to predict it, and the actual prediction performance of the prediction methods is judged by comparing the fit between the prediction curve and the actual load curve.

5.2 Analysis of Test Results

The comparison criterion chosen for this experiment is the degree of fitting between the load prediction curve and the actual load fluctuation curve under different methods, the higher the degree of fitting, the higher the prediction accuracy of the method for electric load, the specific experimental results are shown in Fig. 2. Among them, the thick line part is the actual power system operating load curve.

In contrast, the fit between the predicted load curves and the actual load operation curves under the two conventional methods is smaller, which proves that the load

Fig. 2 Comparison of power system load forecast curves



prediction method proposed in this paper is better than the conventional prediction methods in terms of prediction accuracy.

6 Concluding Remarks

This paper addresses the problem of inefficiency of conventional load forecasting methods for power systems, combines online compression and reconstruction technology to process the historical operation data of power systems, and compresses and reconstructs the operation data through online learning compression and sensing methods. The load prediction algorithm constructed on this basis can effectively improve the data transmission efficiency, thus improving the algorithm operation efficiency. The experimental results show that the prediction algorithm proposed in this paper can achieve accurate prediction of the operating load of the power system within the specified time, and it is feasible to ensure the operation efficiency while taking into account the prediction accuracy.

References

1. Li F, Sun M (2021) EMLP: short-term gas load forecasting based on ensemble multilayer perceptron with adaptive weight correction. *Math Biosci Eng: MBE* 18(2):1590–1608
2. Gomezroero S, Capretz M, Mir S (2021) Transfer learning by similarity centred architecture evolution for multiple residential load forecasting. *Smart Cities* 4(1):217–240
3. Chafi ZS, Afrakhte H (2021) Short-term load forecasting using neural network and particle swarm optimization (PSO) algorithm. *Math Prob Eng* 2021(2):1–10
4. Ngoc TT, Dai LV, Thuyen CM (2021) Support vector regression based on grid search method of hyperparameters for load forecasting. *Acta Polytechnica Hungarica* 18(2):143–158
5. Envelope Z, Xy A, Fzb C (2022) A two-layer SSA-XGBoost-MLR continuous multi-day peak load forecasting method based on hybrid aggregated two-phase decomposition. *Energy Rep* 8:12426–12441
6. Panda SK, Ray P (2021) Analysis and evaluation of two short-term load forecasting techniques. *Int J Emerg Electr Power Syst* 23(2):183–196
7. Sun C, Xu X (2021) Power system transient stability prediction based on deep residual network. *Comput Simul* 2021(5):77–81