

PAPER • OPEN ACCESS

## Research on Power Load Forecasting Based on Random Forest Regression

To cite this article: Na Liu *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **252** 032171

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

# Research on Power Load Forecasting Based on Random Forest Regression

Na Liu, Yanzhu Hu and Xinbo Ai\*

School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China

\*Corresponding author e-mail: axb@bupt.edu.cn

**Abstract.** In recent years, with the development of artificial intelligence, power forecasting based on big data analysis has gradually become intelligent. In order to improve the prediction accuracy and efficiency of the model in dealing with large volume data, this paper combines the compressed sensing algorithm and the random forest regression algorithm. The discrete cosine transform base is used to sparsely represent the data. The original data is restored by solving the  $l_1$  norm optimization problem to achieve the purpose of denoising. And the processed data is used for regression prediction, it achieves great results. The results indicate that the compressed sensing algorithm can retain more details and get better effect compared with the traditional denoising method. Combined with the random forest regression algorithm, the prediction accuracy of the model is improved. This method can be implemented in the trend prediction of time series data such as power load, which has very important practical significance.

## 1. Introduction

With the development of internet technologies and the measurement of energy consumption analysis, more and more energy consumption data is collected in the power database, which is large in quantity and type, high in value. It is the key issue for power big data applications to discover the rules and mine useful information from the large amount of noisy data. Machine learning is an extremely important means of conducting power big data analysis. How to use machine learning algorithms to effectively analyze power big data is of great significance.

In recent years, the prediction of electric energy based on big data analysis has gradually become intelligent. The random forest algorithm is one of the most prominent algorithms. In the application of random forest regression algorithm for load forecasting, Liu Qichen et al. proposed a short-term power load forecasting method to improve the processing ability of algorithm for big data. This method based on Spark platform and parallel random forest regression algorithm [1]. Zhang Bing et al. used the grey relational analysis method to modify the random forest regression model, which improves the prediction performance of the model [2]. Similarly, a random forest prediction method based on grey projection proposed by Wu Xiaoyu enhanced the prediction accuracy and robustness of the model [3]. Huang Nantian et al. adopted a short-term load forecasting feature selection method using random forests, and proved that the random forest model of the optimal predictive feature subset has higher prediction accuracy than the original model [4]. Huang Han et al. used the random forest algorithm to



increase the prediction accuracy of the model by adjusting the number of trees [5]. Dudek compared random forest model with CART, ARIMA, and neural networks to demonstrate that random forest model has higher precision [6]. Lahouar, A. et al. combined the expert character selection and random forests to enhance the generalization ability of the model by adjusting the input vector [7]. González established a model based on categorical regression trees and random forests to predict short-term electricity in a Spanish market, achieved a certain prediction accuracy [8]. Huang Qingping put forward a short-term load forecasting way based wavelet decomposition and random forest to raise the accuracy of prediction [9]. However, the model cannot meet the requirements of prediction accuracy and efficiency when dealing with large volume and high latitude data. Meanwhile, due to the influence of the surrounding environment of the equipment, the recorded data changes are very complicated, and the noise also has a great influence on the prediction accuracy.

Therefore, this paper combines the compressed sensing and the random forest regression algorithm to achieve the purpose of improving the efficiency and predictive ability of the model. On the basis of the ability to recover the original signal, the compressed data is used to denoise the original data. And then the random forest regression algorithm is served as predicting.

## 2. Theoretical knowledge

The text and data from the EIA report are described in natural language. Their matching can be regarded as the process of NLP. Firstly, a corpus which is suitable to EIA report should be established. Then, the subjective and objective weighting method describes the importance and frequency of vocabulary in EIA method. Lastly, the text and data is matched by the regular expression.

### 2.1. Compressed sensing algorithm

Compressed sensing theory involves three core issues: sparse representation, random sampling, and signal recovery.

Compressed sensing theory involves three core issues: sparse representation, random sampling, and signal recovery. In terms of sparse representation, signals can be accurately expressed by basis functions such as Cosine basis and wavelet basis. From the mathematical point of view, for the finite-length signal vector  $X \in \mathbb{R}$ , assuming that  $\varphi = \{\varphi_i\}_{i=1}^n$  is an orthogonal basis of  $\mathbb{R}^n$ , the signal  $X \in \mathbb{R}$  can be linearly represented as:

$$X = \sum_{i=1}^n s_i \varphi_i \quad (1)$$

where,  $s_i$  is the coefficient of  $X$ :  $s_i = \langle X, \varphi_i \rangle$

In this way,  $X$  can be expressed as  $\varphi s$ ,  $\varphi$  is a  $n \times n$  matrix, where  $\varphi_1, \dots, \varphi_n$  is the column vector. If most of the elements of  $s_i$  are 0, the coefficient  $S$  is sparse. If the signal  $X \in \mathbb{R}$  is sparse in the orthogonal basis  $\varphi$ , then the sparse coefficient  $S$  can be expressed as:

$$s = \varphi^T X \quad (2)$$

As for random sampling, define an  $m \times n$  ( $m < n$ ) observation matrix that is not related to the transformation basis  $\varphi$ . And use it to measure the sparse coefficient  $S$  and obtain the observation vector  $Y = \Phi s$ . So the entire perception process is as follows:

$$Y = \Phi \varphi^T X \quad (3)$$

The observation procedure is the process of capturing useful information in the signal and compressing the high-dimensional signal into a low-dimensional observation signal, which must satisfy the finite equidistance property:

$$(1 - \delta_k) \|s\|_2^2 \leq \|A_k s\|_2^2 \leq (1 + \delta_k) \|s\|_2^2 \quad (4)$$

where  $x$  is the sparse signal, and if  $\delta_k < 1$ , the observation matrix is said to satisfy the  $K$ -order finite equidistant property. The RIP property ensures that the observation matrix satisfies the mapping relations from the original space to the sparse space.

In terms of signal recovery, since the observation number  $m$  is much smaller than the signal length  $n$ , it converts to the problem of solving the underdetermined equations  $Y = \Phi \varphi^T X$ . And there are infinite solutions in the equation. If the sparse matrix satisfies the RIP condition, the problem can be converted to a solution problem:

$$\min = \|\varphi^T X\|_0, \quad s.t. Y = \Phi \varphi^T X \quad (5)$$

where  $\|x\|_0$  represents the number of non-zero elements in  $x$ . In this case, it is necessary to list the possible linear combinations of all non-zero positions in  $x$  to obtain the optimal solution. As the calculation of the above formula is unstable, it can be equivalent to solving a simpler  $l_1$  optimization problem that has the same solution. So the question is converted to:

$$\min = \|\varphi^T X\|_1 \quad s.t. Y = \Phi \varphi^T X \quad (6)$$

where  $\|x\|_1$  represents the  $l_1$  norm of the vector  $x$ . This problem becomes a problem of convex optimization, which can be solved by the linear programming. In the practical application of compression sensing, the measurement process may introduce noise. Suppose Gaussian noise is  $g$ , then the formula  $Y = \Phi \varphi^T X$  becomes  $Y = \Phi \varphi^T X + g$ , so it is necessary to rewrite the equality constraint to an inequality constraint, that is:

$$\min = \|\varphi^T X\|_1 \quad s.t. \|\Phi \varphi^T X - Y\|_1 < \varepsilon \quad (7)$$

## 2.2. Random forest regression algorithm

The establishment process of the random forest regression model mainly includes two steps, that is building a decision tree and forming a random forest. The random forest model is a combined model composed of many regression decision trees. The parameters of each decision tree are independent and identical distributed random variables. The regression decision trees give a prediction for each input sample. Its basic idea is to extract  $K$  times sampling from the original input samples by Bootstrap resampling, and obtain  $K$  samples with the same size as the original input samples. Then, establishing a decision tree with these  $K$  sample. The regression prediction results given by each decision tree are averaged to obtain the final prediction result.

From the perspective of mathematical analysis, the random forest regression algorithm is described. The specific steps are as follows [10]:

(1) Firstly,  $k$  training sets  $\theta_1, \theta_2, \dots, \theta_k$  are randomly selected by bootstrap resampling method. The training sets can generate corresponding decision trees  $\{T(x, \theta_1)\}, \{T(x, \theta_2)\} \dots, \{T(x, \theta_k)\}$ , and then these  $K$  decision trees form a random forest.

(2) Assuming that the dimension of the input sample is  $M$ . In the process of node splitting,  $m$  features are randomly extracted from the  $M$ -dimensional features as the split feature set of the node. And the value of the  $m$  is set according to the size of the sample. Generally, the value of  $m$  in the entire random forest formation process remains constant.

(3) No pruning is performed on each decision tree, that is, the depth of the tree is not limited to achieve the maximum growth.

(4) When a new sample  $X$  is inputted, the prediction of each decision tree  $T(\theta)$  can be obtained by averaging the observations of the leaf nodes  $l(x, \theta)$ . If an observation value  $X_i$  belongs to the leaf node  $l(x, \theta)$  and it is not 0, then the weight vector  $\omega_i(x, \theta)$  is:

$$\omega_i(x, \theta) = \frac{\{X_i \in R_l(x, \theta)\}}{\{j: X_j(x, \theta)\}} \quad (8)$$

where the sum of the weights  $\omega_i(x, \theta)$  is 1.

(5) Given a new input sample  $X$ , the predicted value of each decision tree can be obtained by the weighted average of the variable  $Y_i$  ( $i = 1, 2, \dots, n$ ). The predicted value of a single decision tree is obtained according to equation (9):

$$\bar{\mu}(x) = \sum_{i=1}^n \omega_i(x, \theta) Y_i \quad (9)$$

(6) By averaging the decision tree weights  $X = X_i (i \in \{1, 2, \dots, n\})$ , the weight  $\omega_i(x)$  of each observation  $i \in \{1, 2, \dots, n\}$  is obtained as formula (10):

$$\omega_i(x) = k^{-1} \sum_{j=1}^k \omega_i(x, \theta_j) y_j \quad (10)$$

(7) For any  $y$ , the prediction of random forest can be recorded as  $\bar{\mu}$ :

$$\bar{\mu}(x) = \sum_{i=1}^n \omega_i(x, \theta) Y_i \quad (11)$$

### 2.3. Evaluating indicators

The final predicted results are evaluated to test the performance of the model. The error between the predicted value and the real value and the degree of model fitting are usually used to determine the quality of the model.

The common indicators for describing error are *MAE*, *RMSE*, and *MAPE*. The index for measuring the fitting of the model is  $R^2$ . The specific definition is given as following:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| * 100\% \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

where  $\hat{y}_i$  is the predicted value of the input sample,  $y_i$  is the real value of the input sample,  $\bar{y}$  is the mean of the input samples, and  $n$  is the total number of tested samples.

$R^2$  is used to measure the data fitting degree of the model. Its value ranges from 0 to 1. The closer the value is to 1, the better the fitting effect is. At the same time, it can also be negative values. When its value is negative, it means that the model can't fit the data at all. *MAPE* represents the relative error between the predicted value and the real value. Obviously, the smaller the value of *MAPE*, the smaller the error between the predicted value and the real value, the higher the prediction accuracy of the model. In summary, the model performance evaluation indicators selected in this paper are  $R^2$  and *MAPE*.  $R^2$  is used to measure the fitting degree of the model. *MAPE* is used to measure the prediction accuracy of the model.

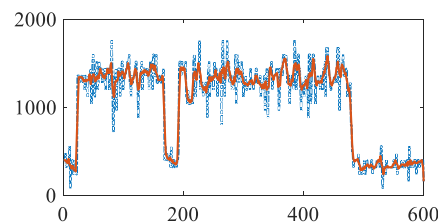
### 3. Experiment

According to the related algorithms mentioned above, the experiment mainly includes two steps, namely data denoising and prediction. First, the data is denoised by the compressed sensing algorithm. Secondly, the random forest regression model is trained to obtain the predicted value  $\hat{X}$  of the original data  $X$ .

The data in this paper is the power monitoring data of an automobile factory. The data includes various information such as weather, temperature, date, week and the 24-hour power consumption of the equipment in each workshop. The length of the training data should be  $2^n$  ( $n>0$ ) in order to create an orthogonal sparse matrix. Therefore, the experiment selects 8874 data from January 1 to December 31 for the study. The first 8192 data is used as the training set to train the model, and the last 592 data is used as test data to test the model.

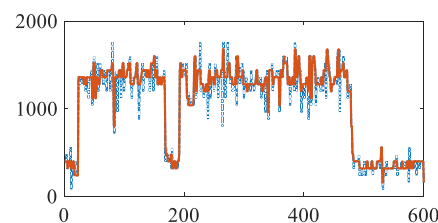
### 3.1. Data process

In the process of compressed sensing denoising, the most important thing is to select the appropriate sparse basis for sparse transformation. In this paper, the discrete cosine transform is used as the sparse basis, and the number of sampling is 210. In order to compare the denoising effect of the compressed sensing method and other traditional algorithms, median filtering and smoothing filtering are selected for comparison. The denoising effect of the three methods is shown in figure 1-3.

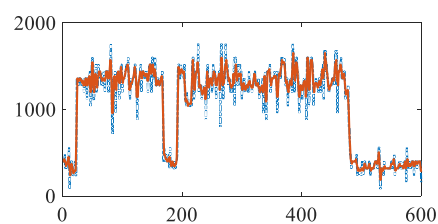


**Figure 1.** Compressed sensing denoising effect

Diagram



**Figure 2.** Median filtering denoising effect diagram



**Figure 3.** Smoothing filter denoising effect diagram

Compare the above figures, it shows that the compressed sensing algorithm can reduce the noise in the data, and retain more details. The denoising effect is better than the other two methods.

Since the data in this paper is the time series, that is, the data of the current day is related to the data of the previous few days. Therefore, in the process of random forest regression prediction, the sliding window prediction strategy is adopted. The one-dimensional time series is transformed to a matrix form. The size of the selected sliding window is 6, that is, the data from the first 5 days are used to predict the 6th day.

After transformation, the form of sample for model training is as follows:

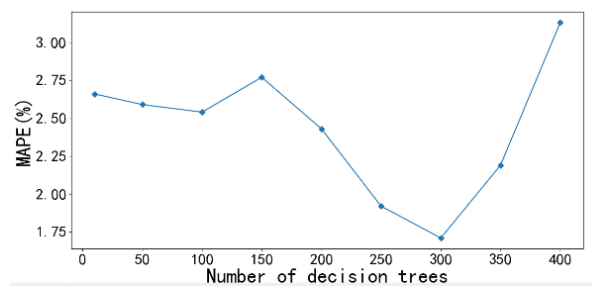
$$X = \begin{bmatrix} x_1 & \cdots & x_5 \\ \vdots & \cdots & \vdots \\ x_{n-5} & \cdots & x_{n-1} \end{bmatrix} \quad Y = \begin{bmatrix} x_6 \\ x_7 \\ \cdots \\ x_n \end{bmatrix}$$

Where  $n$  is the number of training samples, and  $X$  is the input vector containing the parameters such as weather, temperature and power consumption.  $Y$  is the power consumption data of the 6th day.

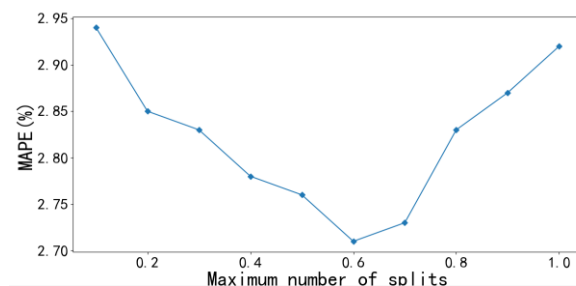
### 3.2. Experiment results

When constructing a random forest regression prediction model, the model parameters to be determined are as follows: number of decision trees, depth of decision trees, and maximum number of divisions. In this section, the pre-processed sample data is used for prediction, and the influence of different model parameters on the prediction effect of the algorithm is compared, and the optimal model parameters are selected.

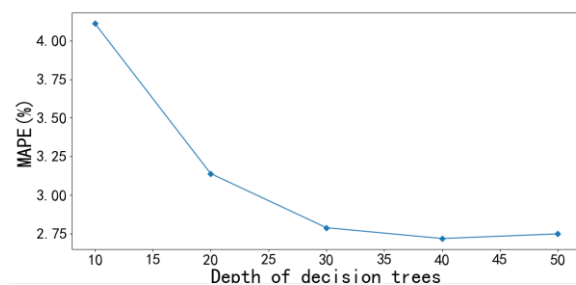
(1) The number of decision trees. In the condition of data set (8192) and other parameters (decision tree depth, split strategy, maximum split number) remain constant, consider the influence of decision trees number on the model. The graph of MAPE changes with the number of decision trees is shown in Figure 4.



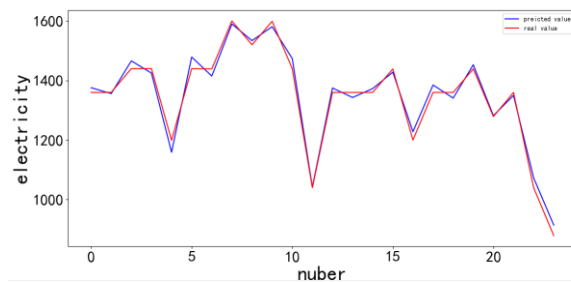
**Figure 4.** The graph of MAPE changes with the number of decision trees



**Figure 5.** The graph of MAPE changes with the maximum number of divisions



**Figure 6.** The graph of MAPE changes with the decision trees depth



**Figure 7.** The prediction of electricity consumption on November 6

The graph indicates that when the number of decision trees is small, the random forest regression error is large and the performance is poor. As the number of decision trees increases, the error between the predicted and real values decreases gradually. But the algorithm becomes more complex and takes more time.

(2) Decision tree depth. In the data set (8192) and other parameters (the number of decision trees, split strategy, maximum split) remain constant, study the impact of decision trees depth on the model. The MAPE trend is shown in Figure 5.

It can be seen from the figure that as the depth increases, the prediction accuracy of the model is gradually improved. When the depth of the decision tree increases to 40, the value of MAPE tends to be flat and stable.

(3) The maximum number of divisions. In the data set (8192) and other parameters (the number of decision trees, split strategy, maximum split) remain constant, consider the effect of the maximum number of splits on the model. The graph of MAPE changes is shown in Figure 6.

The figure shows that the value of MAPE decreases with the increase of the maximum number of splits, and the prediction accuracy of the model generally increases. When the maximum split number is around 0.6, the MAPE reaches the minimum value.

The optimal parameters obtained by the GridSearchCV ( ) function:  $n\_estimators = 300$ ,  $max\_features = 0.5$ ,  $max\_depth = 40$ . The model is used to predict the electricity consumption for 24 hours a day on November 6, and the prediction results are shown in Figure 7.

The evaluation results obtained are as follows:  $R^2$  is 0.985, MAPE is 2.13%.

The comparison between the real value and the predicted value is shown in Table 1.

**Table1.** Relative error between true and predicted values

| Date  | real value | predicted value | Relative error (%) | date  | real value | predicted value | Relative error (%) |
|-------|------------|-----------------|--------------------|-------|------------|-----------------|--------------------|
| 8:30  | 1520       | 1564            | 2.92               | 14:30 | 1360       | 1373            | 0.96               |
| 9:30  | 1599       | 1581            | 1.12               | 15:30 | 1439       | 1428            | 0.76               |
| 10:30 | 1440       | 1474            | 2.36               | 16:30 | 1360       | 1388            | 2.05               |
| 11:30 | 1040       | 1042            | 0.19               | 17:30 | 1360       | 1385            | 1.84               |
| 12:30 | 1360       | 1375            | 1.10               | 18:30 | 1439       | 1420            | 1.32               |
| 13:30 | 1360       | 1343            | 1.25               | 19:30 | 1279       | 1319            | 3.09               |

It can be seen from the above results that the performance of the model is influenced by the number and depth of the decision tree, and the maximum number of splits. Within a certain range, the performance of the model is improved as the value of the parameter increases. However, the increase in the number and depth of trees will also lead to an increase in model training time and model complexity. In this paper, combined with the compressed sensing denoising algorithm, the regression model is constructed. When the number of decision tree is 300, the depth of tree is 40, and the maximum split number is 0.6, the model has the best prediction effect.



#### 4. Conclusion

In this paper, the model of random forest regression algorithm combined compressed sensing achieves to predict the power consumption prediction. The influence of various parameters of random forest on the prediction accuracy of the model is analyzed. The conclusions is following:

(1) The optimal value of the trees number is 250~300. When the number is greater than 300, the complexity and training time of the model will increase and the accuracy of the model decrease.

(2) The optimal depth of the trees is 40. The depth of the decision tree improves the accuracy of the model while increases the complexity of the algorithm. Under certain conditions, this parameter can be increased to improve the prediction accuracy of the model.

(3) The optimal maximum number of splits is 0.6, and the maximum split number can also improve the accuracy of the model. In the case of a large number of sample features, the value can be reasonably took to control the generation time of the decision tree.

In this paper, the short-term prediction of electricity consumption is initially realized, and some practical conclusions are drawn. However, there are still some shortcomings in the work of this paper. Such as only one year of power data is used in the paper, and the established model is only suitable for short-term power load forecasting.

#### Acknowledgments

This work was financially supported by Beijing Municipal Science and Technology Project (No. Z181100009018003).

#### References

- [1] Liu Q, Lei J , Hao J , et al. Short-term power load forecasting based on spark platform and parallel random forest regression algorithm model [J]. Electric Power Construction, 2017.
- [2] Zhang B, Zhou B, Shi M, et al Short-term load forecasting based on grey relational analysis and random forest regression model [J]. Water Resource and Power, 2017 (4):203-207.
- [3] Wu X.Y, et al. Power System Short term Load Forecasting Based on Improved Random Forest with Grey Relation Projection [J]. Automation of Electric Power Systems, 2015, 39 (12):50-55.
- [4] Huang N, Lu G, Xu D. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest [J]. Energies.2016.9 (10):767.
- [5] Huang H, Sun K, Liu D. Hourly Load Forecasting of Power System Based on Random Forest [J]. Shaanxi Electric Power, 2018 (5):8-14
- [6] Dudek G. Short-Term Load Forecasting Using Random Forests [J]. 2015, 323:821-828.
- [7] Lahouar A, Slama J B H. Day-ahead load forecast using random forest and expert input selection [J]. Energy Conversion & Management, 2015, 103 (October 2015):1040-1051.
- [8] González, C, Mira-Mcwilliams J , Juárez, I. Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests [J]. Generation Transmission & Distribution Iet, 2015, 9 (11):1120-1128.
- [9] Huang Q, Li Y, Liu S, et al. Short term load forecasting based on wavelet decomposition and random forest [C]. Proceedings of the Workshop on Smart Internet of Things. ACM, 2017: 2.
- [10] Wang Z, Wang Y, Zeng R, et al. Random Forest based hourly building energy prediction[J]. Energy and Buildings, 2018, 171: 11-25.