

Daily Electricity Consumption Forecasting Based on Lazy Learning

Haiying Li, Bingfang Yang

Department of Electrical Engineering, University of Shanghai for Science and Technology, Yangpu District, Shanghai 200093, China

Abstract. Daily electricity consumption is varying randomly. To improve forecasting accuracy, a Lazy Learning (LL) model is proposed. LL aims to build the regression forecasting models upon vectors which are chosen by K -vector nearest neighbors (K -VNN) method. K -VNN can solve overfitting problem and high accuracy can be ensured. Since there are many factors related to electricity consumption, Grey T's correlation degree is used to determine key indexes to further improve the running efficiency of the model. In addition, fuzzy C-means (FCM) clustering is applied to explore the similar scenarios, then the searching scope of LL is reduced. A case studied in one building in Shanghai shows the proposed method can enhance the accuracy and efficiency of electricity consumption forecasting.

1 Introduction

Accurate forecasting for daily electricity consumption is of great significance to power network planning and user's management decision. The establishment of an efficient forecasting model based on big data of power system is an important topic of electricity consumption forecasting. Time series methods are adopted to identify the future electricity consumption based on historical data [1-3]. Besides these historical electricity consumption data, external environment such as temperature and weather characteristics is closely related to future load, so multivariable coupling methods are put forward [4, 5]. To further improve forecasting accuracy, the integrated methods are proposed [6, 7], in which different weights are used to combine the forecasting results of various algorithms. The literatures above provided important references for load forecasting, but the following issues need to be further researched:

1) The unified parameter models are adopted in present electricity consumption forecasting. However, the unified parameter models incorporate every error, which will result in overfitting problem.

2) The randomness of electricity consumption is gradually increasing. The forecasting accuracy can only be improved by selecting similar electricity consumption scenarios.

For this purpose, a Lazy Learning (LL) method [8, 9] is proposed to forecast electricity consumption. LL algorithm is based on "similar inputs produce similar outputs". Similar scenarios are chosen by K -vector nearest neighbors (K -VNN), so overfitting problem is overcome and high accuracy can be ensured.

To improve the running efficiency, Grey T's correlation degree [10] is used to find key indexes to improve the efficiency of the model. In addition, fuzzy

C-means (FCM) clustering [11] is applied to explore similar scenarios and reduce searching scope of modeling. The results demonstrate that the proposed method can enhance the accuracy and efficiency of electricity consumption forecasting.

2 Lazy learning algorithm

2.1 K -VNN

K -vector nearest neighbors (K -VNN) method is employed to choose similar input data from database for forecasting day f . Input vector of f day $X_f=[x_{f1}, x_{f2}, \dots, x_{fN}]^T$ is given. Here N is the dimension of input vector.

To choose the nearest K neighbors, the distance and angle between X_f and X_t is calculated. X_t is sample t in the database, $t \in T$, T is time set.

The Euclidean distance $d_t(X_t, X_f)$ and angle $\theta_t(X_t, X_f)$ are defined as follows:

$$d_t(X_t, X_f) = \sqrt{\|X_t - X_f\|_2} = \sqrt{\sum_{n=1}^N (x_{tn} - x_{fn})^2} \quad (1)$$

$$\theta_t(X_t, X_f) = \cos^{-1}\left(\frac{X_t^T X_f}{\|X_t\|_2 \cdot \|X_f\|_2}\right) \quad (2)$$

Once $d_t(X_t, X_f)$ and $\theta_t(X_t, X_f)$ are obtained, K nearest neighbors are selected as follows:

1) $\cos(X_t, X_f) < 0$ indicates that X_t and X_f are in opposition, and X_t should be abandoned.

2) While $\cos(X_t, X_f) \geq 0$, $D(X_t, X_f)$ is further used to describe the distance and angle between X_f and his neighbors. The larger $D(X_t, X_f)$ is, the nearer these two vectors are.

$$D(X_t, X_f) = \alpha e^{-d(X_t, X_f)} + (1 - \alpha) \cos(\theta(X_t, X_f)) \quad (3)$$

^a Corresponding author: y-somnus-love@foxmail.com

where α is the weight, $0 \leq \alpha \leq 1$. Logistic function is used to map $d(\mathbf{X}_i, \mathbf{X}_j)$ to the interval of $[0, 1]$, which is consistent with the range of $\cos(\mathbf{X}_i, \mathbf{X}_j)$. Furthermore, K nearest data \mathbf{X}_{jk} ($k=1, 2, \dots, K$) are selected by sorting $D(\mathbf{X}_i, \mathbf{X}_j)$ in descending order. To enhance forecasting accuracy, the proportion of K in historical database samples is adjusted to 10%.

2.2 Prediction model

According to the nearest neighbors $\mathbf{X}_F = \{\mathbf{X}_{f1}, \mathbf{X}_{f2}, \dots, \mathbf{X}_{fK}\}$ and the electricity consumption $\mathbf{Y}_F = [y_{f1}, y_{f2}, \dots, y_{fK}]^T$, the first order linear regression model is established.

$$\hat{\mathbf{Y}}_F = \mathbf{X}_F^T \boldsymbol{\beta} \quad (4)$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_N]$ is the weight of input vector. $\hat{\mathbf{Y}}_F = [\hat{y}_{f1}, \hat{y}_{f2}, \dots, \hat{y}_{fK}]^T$ is the forecasting outputs.

The weighted sum-squared residual is taken as an evaluation criterion for the accuracy of the model.

$$J = \sum_{k=1}^K (y_{fk} - \hat{y}_{fk})^2 D(\mathbf{X}_{fk}, \mathbf{X}_f) \quad (5)$$

where $D(\cdot)$ is the residual weight. The model parameter $\boldsymbol{\beta}$ can be obtained by minimizing J .

The output of \mathbf{X}_f is calculated by formula (6).

$$\hat{y}_f = \mathbf{X}_f^T \boldsymbol{\beta} \quad (6)$$

LL algorithm selects the nearest neighbors to model for each forecasting day, which can not only solve the problem of overfitting but increase forecasting accuracy as well.

3 Data processing

3.1 Data standardization

Data at different scale should be standardized first. Z-score method is adopted for standardization.

$$x_{m,t} = \frac{x_{m,t} - M_n}{S_n} \quad (7)$$

where $x_{m,1}$ is the initial input. $x_{m,t}$ is the standardized input. $n=1, 2, \dots, N$. M_n is the mean value. S_n is the standard deviation. standardized vector at t time is $\mathbf{X}_t = [x_{t1}, x_{t2}, \dots, x_{tN}]^T$.

3.2 Inputs selection

Daily load is related to historical daily load, temperature, weather factors and holiday. To identify these factors which impact daily load obviously, Grey T's correlation degree [8] is used to measure the relation degree.

In Grey T's correlation degree calculation, daily load $[y_1 \ y_2 \ \dots \ y_T]$ is taken as a parent factor, and each input $[x_{1n} \ x_{2n} \ \dots \ x_{Tn}]$ is as a sub factor. Grey correlation degree r_n of the n th factor is calculated as follows:

$$r_n = \frac{1}{T-1} \sum_{t=2}^T \xi_m \quad (8)$$

where

$$\xi_m = \begin{cases} \text{sgn}(\Delta y_t \cdot \Delta x_m) \cdot \frac{\min(|\Delta y_t|, |\Delta x_m|)}{\max(|\Delta y_t|, |\Delta x_m|)} & (\Delta y_t \cdot \Delta x_m) < 0 \\ 0 & (\Delta y_t \cdot \Delta x_m) \geq 0 \end{cases} \quad (9)$$

$$\Delta y_t = \frac{y_t - y_{t-1}}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (10)$$

$$\Delta x_m = \frac{x_{m,t} - x_{(t-1)n}}{\frac{1}{T-1} \sum_{t=2}^T |x_{m,t} - x_{(t-1)n}|} \quad (11)$$

The correlation degree $|r_n| \leq 1$, the greater the $|r_n|$ is, the higher the correlation degree is. $r_n > 0$ indicates that parent factor and sub factor have positive correlation, otherwise have negative correlation.

Considering the trend of temperature influence on daily load in different seasons, the correlation degrees are determined based on four individual seasons.

3.3 Fuzzy c-means clustering

LL algorithm determines K neighbors by tracing the whole database, which will take a long time. Therefore, FCM is used to classify database into small clusters.

To realize the classification, the distance between clusters should be minimized.

$$\min J(\mathbf{U}, \mathbf{c}_1, \dots, \mathbf{c}_C) = \min \sum_{t=1}^T \sum_{i=1}^C u_{ti}^m d_{ti}^2 \quad (12)$$

$$\sum_{i=1}^C u_{ti} = 1, \forall t = 1, \dots, T \quad (13)$$

where C is the number of clusters. \mathbf{c}_i ($i=1, 2, \dots, C$) is the i th cluster center. \mathbf{U} is the membership matrix of $T \times C$, and the element u_{ti} indicates the degree how \mathbf{X}_t belongs to the fuzzy set \mathbf{c}_i , $0 \leq u_{ti} \leq 1$. $d_{ti} = \|\mathbf{X}_t - \mathbf{c}_i\|$ is the Euclidean distance between the t th input vector and the i th cluster center. $m \in [1, \infty]$ is a weighted index, and set to 2 here.

The necessary conditions for equations (12)~(13) are obtained as follows by constructing Lagrangian functions.

$$\mathbf{c}_i = \frac{\sum_{t=1}^T u_{ti}^m \mathbf{X}_t}{\sum_{t=1}^T u_{ti}^m} \quad (14)$$

$$u_{ti} = \frac{1}{\sum_{j=1}^C \left(\frac{d_{ti}^2}{d_{tj}^2} \right)^{2/(m-1)}} \quad (15)$$

The optimal classifications are determined by iteration. The optimal evaluation index is (16). When $L(C-1) < L(C-2)$ and $L(C-1) < L(C)$ are satisfied, $C-1$ is the optimal number of classification.

$$L(C) = \frac{1}{T} \frac{J(\mathbf{U}, \mathbf{c}_1, \dots, \mathbf{c}_C)}{\min \|c_{k1} - c_{k2}\|} \quad (16)$$

4 Case studies

4.1 Selection of input factors

FCM-LL algorithm proposed is used to forecast daily electricity consumption of one building in Shanghai. The data are covered from January 1, 2017 to December 31, 2017.

The candidate inputs include the highest temperature $x_{t1,I}$, the lowest temperature $x_{t2,I}$, weather factor $x_{t3,I}$, holiday $x_{t4,I}$, average temperature $x_{t5,I}$, and daily load in the previous week $x_{t6,I} \sim x_{t12,I}$. For $x_{t4,I}$, 0 represents holidays, and 1 represents weekdays. Weather factors are refined to [0, 1] [12]. For example, 0.6 represents cloudy to moderate rain and 1 represents sunny.

45 days' data are taken from summer, autumn and winter individually for correlation analysis, and correlation degree are denoted by $r_{n,S}$, $r_{n,F}$ and $r_{n,W}$.

Table 1. Grey T's correlation degree between input factors and electricity consumption

	$r_{n,S}$	$r_{n,F}$	$r_{n,W}$
$x_{t1,I}$	0.8383	-0.7401	-0.7247
$x_{t2,I}$	0.7206	-0.5936	-0.6515
$x_{t3,I}$	0.6638	0.4037	0.5034
$x_{t4,I}$	0.7011	0.4087	0.6221
$x_{t5,I}$	0.4865	-0.4311	-0.2196
$x_{t6,I}$	-0.4126	0.4468	-0.4340
$x_{t7,I}$	0.2551	-0.3996	0.5100
$x_{t8,I}$	0.4021	0.3121	0.3826
$x_{t9,I}$	0.3846	0.3844	0.2679
$x_{t10,I}$	0.2371	0.3675	-0.3502
$x_{t11,I}$	-0.2478	0.3511	0.1478
$x_{t12,I}$	0.1892	0.3322	0.2698

As shown in Table 1, time series of daily electricity consumption with future load has a weak correlation and cannot be used as inputs. In addition, daily electricity consumption is related to the highest temperature, the lowest temperature, holiday and weather factors. Therefore, a 365×5 database ($X_{t,I}, y_t$) is set up.

4.2 Data clustering

To further reduce the search scope of LL algorithm, the number of clustering data is limited to 50, and Table 2 is an example.

This are 20 data in this cluster, whose clustering center is $c = \{21.14, 11.53, 0.71, 0.85, 320.93\}$. The data in this class has similar characteristics.

The 16th data has 21.93% relative offset from the class, which is judged as bad data.

Table 2. FCM example

	$x_{t1,I}/^{\circ}\text{C}$	$x_{t2,I}/^{\circ}\text{C}$	$x_{t3,I}$	$x_{t4,I}$	$y_t/(\text{kW}\cdot\text{h})$
1	19	15	0.6	0	323.10
2	17	13	0.5	1	324.15
3	22	12	0.8	1	325.98
4	22	11	1	0	316.63

5	23	14	0.9	0	321.41
6	20	11	0.9	1	327.49
7	19	13	0.8	1	309.19
8	23	13	1	1	306.02
9	23	11	1	0	308.39
10	20	8	1	0	330.26
11	17	6	1	1	318.01
12	19	9	1	1	304.31
13	22	9	1	1	310.53
14	23	10	1	1	320.18
15	20	7	1	0	318.42
16	19	11	0.8	1	391.41
17	22	10	0.9	1	319.04
18	23	11	1	1	320.94
19	23	12	0.9	1	334.97
20	17	9	0.7	0	329.63

4.3 Algorithm performance

Algorithm performance is described by E_{MAPE} and E . E_{MAPE} is mean absolute percentage error. Error E indicates the ratio between the data with e_i less than the preset error and total forecast days m .

$$E_{\text{MAPE}} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (17)$$

$$e_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (18)$$

To evaluate the algorithm performance, traditional LL algorithm with weather factors (model I), FCM-LL algorithm without weather factors (model II), and FCM-LL algorithm with weather factors (model III) are selected.

Leave-one-out cross validation and sampling validation are adopted to verify the performance of the algorithm. Leave-one-out cross validation selects one sample in turn. Sampling validation selects 20 data randomly.

Performance parameters of these three models are shown in Table 3, where t_{F_i} and t_{T_e} are the test times.

As shown in Table 3, weather factors improve the accuracy of forecasting. Compared with the traditional model I, model III has higher forecasting accuracy and shorter running time.

Figure 1 further demonstrates that model III can track the daily load, and can meet the needs of practical engineering.

Table 3. Comparison of algorithm results

Performance parameters	Model I	Model II	Model III
t_{T_e}/s	0.4537	0.3471	0.3963
Leave-one-out cross validation			
$E_{\text{MAPE}}/\%$	5.492	5.631	4.375
$E < 5\%$	59.93	66.73	72.6
$E < 15\%$	88.61	87.59	98.5

	t_{Fi}/s	0.0137	0.0067	0.0064
Sampling validation	$E_{MAPE}/\%$	5.461	5.063	4.362
	$E<5\%$	72.83	68.67	85.38
	$E<15\%$	93.64	94.32	100

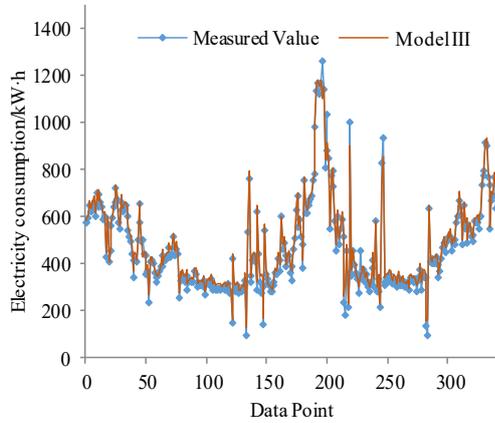


Figure 1. Forecasted values versus measured values

4.4 Error analysis

To analyze the error differences among three models, electricity consumption of November 6, 2017 is taken as an example, the selected modeling data are shown in Table 4.

Table 4. Modeling based data

	$x_{1,i}/^{\circ}C$	$x_{2,i}/^{\circ}C$	$x_{3,i}$	$x_{4,i}$	$y_i/(kW\cdot h)$
Test point	19	12	0.8	1	311.41
	19	13	0.8	1	309.19
Model I	22	12	0.8	1	325.98
	19	11	0.8	1	391.41
Model II	19	13		1	309.19
	17	14		1	436.48
	17	13		1	324.15
Model III	19	13	0.8	1	309.19
	22	12	0.8	1	325.98
	22	11	0.8	1	319.87

As Table 4 shows, more can be mined and summarized as follows:

1) K -VNN can effectively select neighbors to model for forecasting day, and weather factors have important impacts for clustering and selecting neighbors. $\{17,14,1,436.48\}$ and $\{17,13,1,324.15\}$ cannot be identified in Model II without weather factors which leads to the extral error of Model II.

2) In Model III, data clustering and bad data identification can filter noise data. The database is reduced and the credibility of the data is improved. For example, $\{19,11,0.8,1,391.41\}$ in model I is filtered as bad data and is not included in model III to ensure accuracy.

5 Conclusions

In summary, FCM-LL model proposed in this paper is an effective method to forecasting electricity consumption. The following conclusions can be obtained through case studies:

- 1) Grey T's correlation degree can select effective input vectors.
- 2) FCM algorithm can obviously reduce the researching scale, which can improve running efficiency.
- 3) FCM-LL forecasting modeled by select similar scenarios, so the problem of overfitting is overcome, and forecasting accuracy is further improved.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 51777126, and in part by Key Laboratory of Mining Electrical Equipment and Intelligent Control (Taiyuan University of Technology), Shanxi, China (MEI201601), which are gratefully acknowledged.

References

1. Li G D, Wang C H, Masuda S, et al, *International Journal of Electrical Power & Energy Systems*, 4, 33(2011)
2. Wan Kun, Liu Ruiyu, *Power System Technology*, 11, 36(2012) (In Chinese)
3. Zhang Shuqing, Shi Rongyan, Dong Yulan, et al, *Proceedings of the CSEE*, 22, 35(2015) (In Chinese)
4. Zhao Feng, Sun Bo, Zhang Chenghui, *Proceedings of the CSEE*, 2, 36(2016) (In Chinese)
5. Jiang Min, Gu Dongjian, Kong Jun, et al, *Power System Technology*, 1, 38(2014) (In Chinese)
6. Che J, Wang J, Wang G, *Energy*, 1, 37(2012)
7. Nowotarski J, Liu B, Weron R, et al, *Energy*,98(2016)
8. Cherman E A, Spolaôr N, Valverde-Rebaza J, et al, *Journal of Intelligent & Robotic Systems*, 1, 80(2015)
9. Reyes O, Morell C, Ventura S, *Information Sciences*, 340–341(2016)
10. Huang Hao, Wang Haohua, *Mathematics in Practice and Theory*, 1, 46(2016) (In Chinese)
11. Pal N R, Bezdek J C, *IEEE Transactions on Fuzzy Systems*, 3, 3(2002)
12. Cheng Qiyun, Chongqing University, 2014 (In Chinese)