Hourly electric load forecasting for buildings using hybrid intelligent modelling

Yuanyuan Chen¹, Peiyong Duan¹* and Junqing Li¹

¹School of Information Science and Engineering, Shandong Normal University, Jinan, 250358, China
*Corresponding author’s e-mail: duanpeiyong@sdnu.edu.cn

Abstract. Because of the rapidly increasing total electric load of buildings, effective electric load management should be achieved quickly. This can be realized via electric load forecasting. In this study, a novel clustering-based hybrid prediction model is proposed to predict the 24-daily electric load of buildings. In this study, fuzzy c-means (FCM) clustering, ensemble empirical mode decomposition (EEMD), and some intelligent prediction algorithms are combined. FCM is used to extract the daily data exhibiting similar features, whereas EEMD is used for breaking down the optimal prediction algorithm is selected for each component, and the prediction results are integrated. When compared with the remaining conventional prediction models based on real data, the proposed hybrid model exhibits higher prediction accuracy.

1. Introduction

The total building energy consumption is increasing rapidly because of the acceleration of global industrialization and urbanization. The building energy consumption of China accounts for approximately 28% of its total energy consumption [1]. The buildings in Europe are reported to consume more than 40% of the total electricity load in the region [2]. Thus, the proportion of building load is considerably large; therefore, appropriate methods and strategies should be developed to manage it effectively. Several studies have investigated methods of improving the energy performance of buildings to alleviate environmental problems [3-4]. The International Energy Agency has considered building energy efficiency as one of the five measures to ensure long-term decarburization [5]. In addition to environmental benefits, effective building energy management can also result in massive economic benefits. Buildings with efficient energy management systems and strategies have low operating costs [6]. Therefore, building energy management is very important for achieving global sustainable development. In the previous 25 years, researchers have been developing methods to improve building energy consumption. The prediction of electrical load in buildings has been witnessing huge research interest in recent years, and the results obtained from such studies can be used to formulate various strategies for improving the building energy performance [7]. Complex multi-system energy scheduling and management are performed based on accurate prediction energy consumption [8].

The prediction accuracy is an important factor associated with the development of a load forecasting model. The relevant factors affecting the electric load of buildings will differ based on the operation modes, longitudes, and latitudes of the buildings. If any of these factors are changed, the electric load curve of the concerned building may change. Although many current prediction methods, such as artificial neural networks (ANNs) and support vector regression (SVR) can be used to predict the load,
the prediction accuracy obtained using these algorithms is often not high. The building load data are usually complex owing to the various types of factors involved. Therefore, it is necessary to find a suitable method for handling this problem. Thus, a suitable and highly accurate prediction model should be developed to predict the electric load of residential buildings on an hourly basis. In this study, fuzzy c-means (FCM) and ensemble empirical mode decomposition (EEMD) were used to process the energy consumption data; then, the best-performing prediction algorithm is selected for each component of the energy data, which is followed by data integration.

2. Literature review
Currently, building electric load forecasting models have been considered as the premise of building operation and control. From the modeling perspective, the building electric load models can be classified as white box modeling, black box modeling, and gray box modeling. In a previous study [9], the three models were compared and analyzed, and the results showed that the black box model exhibits high accuracy. The learning algorithms used to implement the black box models include ANN, SVR, decision tree, and statistical methods. In a previous study [9], ANN and SVR were reported as the optimal algorithms for predicting the building load. Recently, several machine learning algorithms have been combined to develop a hybrid modeling algorithm, thereby improving the accuracy of the developed models. Previously [10], the Levenberg–Marquardt algorithm was combined with the OWO–Newton algorithm, in which the best parameter range was 0.87–0.91. In a previous study [11], two ANN-based energy prediction models were proposed for estimating the electric load of domestic hot water heating in Canada. Singular spectrum analysis, SVM, and Cuckoo search algorithm were combined to successfully forecast the electricity load [12]. Furthermore [13], a regional hybrid short-term load forecasting model was proposed, in which SVR was combined with the grasshopper optimization algorithm (GOA). In addition, a hybrid model based on a multi-objective flower pollination algorithm was proposed [14]. A multi-objective flower pollination algorithm was used for optimizing the weight of the prediction models. In a previous study [15], a generalized regression neural network based on the multi-objective firefly algorithm was proposed. A previous study proposed training an SVM model using a dynamic time-warping-based data selection method on related days [16]. Further, a multilayer recursive structure based on the standard back-propagation learning algorithm was adopted [17]. Another study [18] presented a multi-objective optimization model based on the genetic algorithm and ANN. The use of the ANN to predict building loads has been studied [19, 20]. The usage of the differential evolution algorithm was proposed to determine the weight of SVR [21]. The accuracy of the hybrid model is higher compared with that of a single algorithm when the k-means algorithm is combined with the ANN and least squares SVR (LSSVR) [22].

The building load data are complex and changeable because of which many scholars often choose some algorithms to process the data during the initial stage. EEMD is one such algorithm. EEMD is used in many fields to improve the model performance. In a previous study [23], EEMD was combined with a prediction method to improve the prediction accuracy. Another study [24] combined EEMD with a forecasting method based on optimal virtual forecasting to improve the forecasting performance.

Accurate building load prediction can significantly improve electricity consumption. Hilton Worldwide has developed a LightStay platform to forecast future electricity use by tracking both historical load data and weather data and performing corresponding actions to improve the electricity use [25].

3. Hybrid intelligent modelling
This study intends to provide a comprehensive and highly accurate building load prediction model for predicting the next day’s 24-h energy consumption. The building load prediction model in this study is mainly divided into three steps. The first step is clustering. FCM is used to extract similar days’ data. The second step is to use EEMD to deal with the energy consumption data (hourly load in the
clustering results) and obtain the m intrinsic mode function (IMF) and the 1 residual (RES). The third step is to select the best prediction algorithm, including long short-term memory (LSTM), back propagation (BP), and least squares SVR (LSSVR), for each IMF and RES. Finally, the results are integrated. Figure 1 shows the structure of the proposed hybrid model.

![Figure 1. Structure of the proposed hybrid model](image)

3.1. Clustering of daily data with similar features

Before applying the prediction algorithm, the accuracy of the prediction results is affected by the clustering of the predicted building load data. FCM, which is a widely used clustering method, is used herein. It has witnessed remarkable theoretical research efforts and has a deep mathematical basis; its objective function has been successfully applied in many fields. FCM can decide to which group a given data sample belongs based on the membership degree. The inputs of FCM are the average daily temperature, average daily load, and average daily humidity.

\[ x_j \in (1, 2, 3, ..., n) \]

FCM divides \[ x_j \] into \( c \) fuzzy groups; further, it obtains the clustering centers of each group for minimizing the value function of non-similar indexes. The membership degree range of each point is \([0,1]\), and the degree of belonging to each group is determined based on the membership degree. The sum of the membership degree of a data sample is equal to 1. FCM requires two parameters, i.e., the number of clusters \( c \) and the membership degree \( m \). The range of \( c \) is greater than 1 but less than the total number of clustering samples. Herein, \( m \) is set to the empirical value, i.e., 2 [26].

3.2. EEMD of the building load data

Empirical mode decomposition (EMD) is used for analyzing the non-stationary signals, which can be decomposed into components having different wave characteristics (referred to as IMF components). However, EMD has some disadvantages, which cause the inaccuracy of the IMF components. To solve this problem, the EEMD method is proposed.

White noise is added into the signal to be analyzed by EEMD. The spectrum of white noise is evenly distributed so that signals of different timescales can be automatically separated into the corresponding reference scale.

The steps of the EEMD algorithm are as follows.

Step 1: Add normal white noise to the original signal.
Step 2: Consider the signal with white noise as a whole and perform EMD decomposition to obtain each IMF component.
Step 3: Repeat steps 1 and 2, and add a new normal distribution white noise sequence each time.
Step 4: Average the IMF obtained each time as the final result.
3.3. Building load data prediction algorithms

The prediction ability of the prediction algorithm differs in case of different data. Three prediction algorithms were selected to predict each IMF and the RES to achieve high accuracy with respect to building data prediction. Each prediction algorithm has its own characteristics and produces different prediction results. Figure 2 shows the training process of the load data.

Circulating neural network (RNN) is a time series prediction algorithm, but when the predicted sequence is too long, it will produce gradient explosion or gradient disappearance. LSTM is designed to solve this problem. LSTM adds three doors: forgetting door, input door and output door. LSTM has three inputs: the input value at the current time, the output value at the previous time and the state value at the previous time. Two outputs: the output value at the current time and the state value at the current time.

The LSSVR model is an extension of the SVR theory. The LSSVR model is a kernel-function-learning machine based on the principle of structural risk minimization. The original inequality constraint of SVR is replaced by an equality constraint. The quadratic programming problem can be solved using linear equations, improving the accuracy and speed of convergence.

The BP neural algorithm includes two processes, i.e., the forward propagation of signals and back propagation of errors. During the forward propagation of signals, the input signal acts on the output node through the hidden layer, generating the output signal through nonlinear transformation. The back propagation of error is performed to invert the output error through the hidden layer to the input in a layer-by-layer manner.

4. Experience and discussion

4.1. Data

In this study, the load data and relevant weather data of a hotel building in Jinan City, China, are considered to verify the accuracy of the proposed hybrid model. The hotel building, with a five-star
rating, has an area of 67,600 m2. The electric load and weather-related data were collected in real time. The relevant building data used in this study were collected from April 9, 2018, to July 17, 2018.

4.2. Result of clustering
The clustering method employed has already been discussed in detail in Section 3.1. The hotel buildings are ranked from April 9, 2018, to July 17, 2018 (1… 100) and divided into four clusters; the results are presented in Table 1.

| No. | Day ID | Feature                      |
|-----|--------|------------------------------|
| I   | 1-3,8-11,15,,16,18-20,24-26,29-33,36,39,40,43-47,50,51,53,57,64 | Low temperature, working day |
| II  | 13,41,48,49,55,56,62,69,70,71,76,77,83,84,90,91,91,98      | High temperature, rest day |
| III | 6-7,14,21-23,27-28,34-35,42,63                          | Low temperature, rest day |
| IV  | 4,5,12,17,37,38,52-54,58-61,65-68,72-75,85-89,92-96,99,100 | High temperature, working day |

Two groups of data (II and III) are intuitively visualized to express the clustering results more clearly in Figure 3. In group (II), when the power consumption is between 0 and 5 hours, it fluctuates around 300 kWh, while the hourly power consumption in group (III) is less than 300 kWh. The hourly electricity consumption of the two groups increased sharply between 5 and 10 hours. The electricity consumption of group (II) is obviously higher than that of group (III). During the period of 10-15h, the peak hourly building electricity consumption of group (II) fluctuated up and down at 760 kWh, while the peak hourly building electricity consumption of group (III) was below 700 kWh. The value of group (II) was significantly higher than that of group (III). At the low peak 16h, it was observed that the power consumption value of group (II) was higher than 500 kWh; while the low peak of group (III) was lower than 500 kWh. At about 19h, the value of group (II) is 680-800 kWh, while the value of group (III) is 600-700 kWh.

4.3. Function of EEMD
The objective of using EEMD is to handle the nonlinear problem associated with the building data. EEMD was performed at 8:00 with respect to group (I), as shown in Figure 4, where the signal denotes the initial hourly load data. When compared with the original data signal, the regularity of the IMF and RES components improved significantly.
4.4. Selection of the best algorithm

In this study, the best prediction algorithm is selected for each small class of data to improve the prediction accuracy of the hybrid model. This implies that the next day’s 24-h data cannot be predicted only by establishing a general model or 24 models and that more than 24 considerably refined models are required. During the operation process, the three models are applied by considering the changes in the hidden layer and the related internal parameters.

MSE is used to select the best-performing algorithm. Table 2 shows the best prediction algorithm obtained for each type of data when No. = II.

Table 2. Group (II) Selection of Prediction Algorithms

| No. = II | Hour | IMF1 | IMF2 | IMF3 | RES | Hour | IMF1 | IMF2 | IMF3 | RES |
|---------|------|------|------|------|-----|------|------|------|------|-----|
| 1:00    | LSSVR| LSSVR| LSSVR| LSSVR| 13:00| LSSVR| LSSVR| LSSVR| LSSVR|     |
| 2:00    | LSSVR| LSSVR| BP   | BP   | 14:00| BP   | BP   | BP   | BP   |     |
| 3:00    | BP   | BP   | LSTM | LSSVR| 15:00| LSTM | LSSVR| LSTM | BP   |     |
| 4:00    | BP   | BP   | LSSVR| LSTM | 16:00| LSSVR| BP   | LSTM | BP   |     |
| 5:00    | LSSVR| LSSVR| LSTM | SSVR | 17:00| LSTM | BP   | BP   | BP   |     |
| 6:00    | BP   | LSSVR| LSSVR| BP   | 18:00| BP   | BP   | BP   | BP   |     |
| 7:00    | LSSVR| BP   | BP   | BP   | 19:00| BP   | LSSVR| LSSVR| BP   |     |
| 8:00    | LSSVR| LSSVR| LSSVR| LSSVR| 20:00| LSSVR| LSSVR| LSSVR| BP   |     |
| 9:00    | LSTM | LSTM | LSSVR| BP   | 21:00| LSTM | LSTM | LSSVR| LSTM |     |
| 10:00   | BP   | BP   | LSSVR| BP   | 22:00| BP   | BP   | LSSVR| BP   |     |
| 11:00   | BP   | LSSVR| LSSVR| BP   | 23:00| LSSVR| LSSVR| LSSVR| BP   |     |
| 12:00   | LSSVR| CMAC | BP   | BP   | 24:00| BP   | BP   | LSSVR| BP   |     |

4.5. Prediction results and verification

The proposed hybrid model was verified based on the data obtained on May 13, June 18, and July 5. These three days were not randomly selected; these days exhibited different power load characteristics, representing three different electric load operation modes in case of the hotel. June 18 was a national
legal holiday and the number of people going outside to play increased substantially. May 13 was a regular weekend, and the number of residents increased; however, this increase was not obvious compared with that observed on June 18. July 5 was a normal working day.

Figure 5 shows the comparison of the results predicted by the mixed model of each component with the original data collected on June 18. The abscissa is the 24 h of the day, and the ordinate is the load value. Four graphs correspond to the four components obtained after EEMD.

The autoregressive integrated moving average (ARIMA), Elman neural network, and beetle antennae search (BAS–Elman) were used to perform simultaneous forecasting to verify the validity of the proposed hybrid model. ARIMA is a prediction method based on time series, and the BAS–Elman method uses the BAS algorithm to determine the optimal initial weights and thresholds associated with the Elman neural network [27]. Figures 6–8 show the comparison of the three-day prediction results. Tables 3–5 compare the errors of the selected three days in terms of MAPE, MAE, and RMSE and indicate the high accuracy of the proposed hybrid model in predicting the 24-h load.

![Figure 5. Forecast results of each component on June 18](image-url)
Figure 6. Forecast results obtained using three models on May 13

Table 3. Comparison of the errors obtained using three models on May 13

| May 13   | MAPE  | MAE   | RMSE  |
|----------|-------|-------|-------|
| Hybrid Model | 1.1137 | 5.4883 | 7.9756 |
| ARIMA Model    | 11.7036 | 56.9404 | 71.8584 |
| BAS–Elman Model | 5.7172  | 28.8353 | 40.3122 |

Figure 7. Forecast results obtained using three models on June 18

Table 4. Comparison of the errors obtained using three models on June 18

| June 18    | MAPE  | MAE   | RMSE  |
|------------|-------|-------|-------|
| Hybrid Model | 1.0243 | 5.1992 | 6.6886 |
| ARIMA Model    | 14.6632 | 65.9846 | 78.7888 |
| BAS–Elman Model | 5.2164  | 27.1654 | 35.1682 |
5. Conclusions

In this study, various algorithms, such as FCM, EEMD, BP, LSTM, and LSSVR, were used in collaboration to develop a hybrid model for predicting the electric load of a building in the next 24 h to improve the accuracy of the forecasting models. In the proposed hybrid model, FCM is used for data extraction, EEMD is used for data decomposition, and BP, LSTM, and LSSVR are used for data prediction. In the case of the load forecasting tasks, these three forecasting algorithms were used; further, some internal parameters were adjusted, and the number of hidden layers was varied. Based on the observation and comparison of Tables 3–5, the prediction accuracy of the proposed hybrid model was confirmed to be the best, followed by that of the BAS–Elman model; the ARIMA model exhibited the worst prediction accuracy.

Further, ARIMA (9, 1, 10), ARIMA (10, 1, 10), ARIMA (10, 1, 8) were predicted for May 13, June 18, and July 5, respectively. The BAS–Elman model comprised six hidden layers.

The establishment of an electric load prediction model can be considered to be a building energy conservation method. By improving the accuracy of electric load prediction, the future power load changes and related influencing factors can be understood, based on which meaningful energy conservation strategies can be formulated.

References

[1] Umberto, B. (2016) A cross country comparison of building energy consumption and their trends. Resources Conservation and Recycling.

[2] Balaras, C. A., et al. (2007) European residential buildings and empirical assessment of the Hellenic building stock, energy consumption, emissions and potential energy savings. Building and Environment, 42:1298–1314.

[3] Park, S., Lee, J.S, et al. (2016) Estimating thermal performance and energy saving potential of residential buildings using utility bills. Energy and Buildings, 23–30.

[4] Tsanas, A., Xifara, A. (2015) Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 2012, 49:560–567.
[5] IEA. Energy and climate change.
[6] Deb, C., Zhang, F., Yang, J., et al. (2017) A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews, 74:902–924.
[7] Fan, C., Xiao, F., Wang, S., et al. (2014) Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127: 1–10.
[8] Nosratabadi, S.M., Hooshmand, R.A., et al. (2017) A comprehensive review on microgrid and virtual power plant concepts employed for distributed energy resources scheduling in power systems. Renewable and Sustainable Energy Reviews, 67:341–363.
[9] Wei, Y., Zhang, X., Shi, Y., Xia, L., et al. (2018) A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82:1027–1047.
[10] Biswas, M.A.R., Robinson, M.D., N.Fumo, N. (2016) Prediction of residential building energy consumption: A neural network approach. Energy, 117:84–92.
[11] Aydinalp, M., Ugursal, V.I., Fung, A.S. (2004) Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Applied Energy, 79:159–178.
[12] Zhang, X., Wang, J., Zhang, K. (2017) Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electric Power Systems Research. 146:270–285.
[13] Barman, M., Choudhury, N.B.D., Sutradhar, S. (2018) A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy. 145 :710–720.
[14] Xiao, L., Shao, W., Yu, M., et al. (2017) Research and application of a combined model based on multi-objective optimization for electrical load forecasting. Energy. 119:1057–1074.
[15] Xiao, L., Shao, W., Wang, C., et al. (2016) Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. Applied Energy, 180:213–233.
[16] Paudel, S., Nguyen, P.H., Kling, W.L., et al. (2015) Support vector machine in prediction of building energy demand using pseudo dynamic approach. Computer Science.
[17] Kalogirou, S.A., Bojic, M. (2000) Artificial neural networks for the prediction of the energy consumption of a passive solar building. Energy.25:479–491.
[18] Asadi, E., Silva, M., et al. (2014) Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. Energy and Buildings. 81:444–456.
[19] Neto, A.H., Flávio Augusto Sanzovo Fiorelli. (2008) Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy & Buildings.40:2169–2176.
[20] Muralitharan, K., Sakhthivel, R., Vishnuvarthan, R. (2018) Neural network based optimization approach for energy demand prediction in smart grid. Neurocomputing.273:199–208.
[21] Zhang, F., C.Deb, C., Lee, S.E., Yang, J., Shah, K.W. (2016) Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. Energy Build, 126:94–103.
[22] Castelletiroca,,J., Gomezgonzalez,J.F., Calvorolle,J.L., et al. (2019) Short-term energy demand forecast in hotels using hybrid intelligent modeling. Sensors, 19:2485.
[23] Wang, W.C., Xu, D.M., Chau, K.W. (2013) Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. Journal of Hydroinformatics, 15: 1377–1390.
[24] Liu, K., Zhang, Y., Qin, L.(2016) A novel combined forecasting model for short-term wind power based on ensemble empirical mode decomposition and optimal virtual prediction. Journal of Renewable & Sustainable Energy, 8.
[25] Hilton Worldwide. (2018)Energy.
[26] Panapakidis, I.P., (2016) Clustering based day-ahead and hour-ahead bus load forecasting models.
International Journal of Electrical Power & Energy Systems. 80:171–178.

[27] Chen, X., Li, J., Han, Y., et al. (2019) Load Forecasting Based On Elman Neural Network Optimized By Beetle Antennae Search Optimization. In: 2019 1st International Conference on Industrial Artificial Intelligence. Shenyang.