INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES

OPEN ACCESS JOURNAL

Vol 14, Issue 1, (2024) E-ISSN: 2225-8329

Fusion AR-EEMD and Moving Average Model based on Long-Short Term Memory Network for Forecasting Electricity Loads

Liang-Ying Wei

Department of Information Management, Yuanpei University of Medical Technology, Hsin-Chu, Taiwan

Corresponding Author Email: lywei@mail.ypu.edu.tw

To Link this Article: http://dx.doi.org/10.6007/IJARAFMS/v14-i1/20439 DOI:10.6007/IJARAFMS/v14-i1/20439

Published Online: 05 January 2024

Abstract

Economic and industry growth increases the demand for electricity, forecasting electricity loads is critical for distribution and production of electricity. However, the former forecasting models have three shortcomings: (1) Most of the models use single input feature for forecasting and the forecasting performances of those models are not good enough; (2) Past forecasting methods have used EMD (empirical mode decomposition) method to decompose raw time series data containing noise to improve prediction performance, however EMD method can not handle mode mixing problem; and (3) The recurrent neural network encounters the problem of gradient disappearance or explosion when learning the time series of long-term dependencies. Therefore, proposed model uses autoregressive method and use the EEMD (ensemble empirical mode decomposition) algorithm (which can overcome the mode mixing problem) for time series data decomposition. Further, values produced by moving averages are as input attributes of proposed model to improve forecasting performance. Finally, long-short term memory neural networks (which can solve the shortcomings of recurrent neural networks) are used to build predictive models to enhance electricity loads forecasting model. In the last step of this study, practical electricity loads datasets will be collected to validate the proposed model. Experimental results show that proposed model can improve the accuracy of prediction.

Keyword: Ensemble Empirical Mode Decomposition, Long-Short Term Memory Neural Networks, Moving Average, Autoregressive

Introduction

Due to the non-storable nature of electrical energy, the demand for electricity supply must be balanced at all times to ensure that the electricity generated is equal to the demand electricity. In addition, the gap between power production and consumption will have a great impact on the stability of power supply. Especially when a power outage occurs, it will cause factories to shut down and make people's lives inconvenient. In order to ensure that power supply meets demand, accurate power load forecasting models are very needed forecasting tools for short-term power load frequency control, daily power generation demand planning, power dispatching, mid-term maintenance planning and long-term power system expansion plans.

Electricity demand reflects the economic and social activities of the country. Economic growth provides people with opportunities for high-quality life, but it will also continue to increase the demand for electricity. In order to meet the demand of users, the government must continue to build more power plants, or looking for alternative energy sources, or formulating policy plans for sustainable energy development. Therefore, predicting power demand has become an important topic in energy system planning and operation. Generally speaking, power load is affected by the following factors: (1) Time: There are two load peaks each year: summer and winter, and the power load changes greatly within a day; (2) Region: Due to differences in consumption structure, electric load will also vary between different regions; (3) Temperature: Under different climate conditions, extreme temperatures can increase the demand for electricity. At low temperatures, this demand rises because a large amount of electricity is used to supply heating. At high temperatures, a large amount of electricity will be consumed to power cooling equipment.

Most researchers first consider traditional time series models as forecast models for predicting power demand, and many time series models have been proposed and used to deal with different forecast problem (Bollerslev, 1986, Engle, 1982, Huarng, 2001, Song & Chissom, 1993). Engle (1982) proposed the ARCH (Autoregressive Conditional Heteroskedasticity) model, which has been used by many financial and economic analysts, while the GARCH (1986) (generalized ARCH) model is a generalized form of ARCH. Box and Jenkins (1976) proposed the autoregressive moving average (ARMA) model, which combines the moving average process with the linear difference equation to obtain the autoregressive moving average model. The ARMA model predicts under linear stationary conditions. A model describing this uniform non-stationary behavior can be obtained by assuming some appropriate process differences. Therefore, in order to deal with non-stationary data sets, the autoregressive integrated moving average model (ARIMA) (Box & Jenkins, 1976) was proposed, assuming a linear relationship between variables. From the above literature, AR (autoregression) is a basic and important method of time series models. Not all models can be applied to all data sets. The reason is that the application of traditional time series models needs to meet statistical assumptions (Wei, 2013) and linear traditional time series methods cannot be applied to nonlinear power demand data sets. Additionally, most traditional time series models use a single input attribute for forecasting. Past research has shown that using more input attributes related to prediction will help increase the accuracy of prediction(Shiu et al., 2017). In addition, there is a lot of noise in the original power forecast data, which is caused by weather factors and different environmental conditions. The performance of traditional time series models is poor because these models use raw data containing noise (Wei, 2016).

The moving average method is a common method that uses a set of recent actual data values to predict the trend of time series values in one or more future periods. The moving average method is suitable for spot prediction. In addition, the moving average method can effectively eliminate random fluctuations in forecasting, the moving average method is a simple smoothing forecasting technology. Based on time series data, the sequential average value containing a certain number of items is calculated sequentially to reflect the long-term trend. The moving average method has been applied in many different fields and has excellent predictive capabilities (Bayer et al., 2017, Zhang et al., 2016). Therefore, this study combines the moving average method with the prediction model.

In recent years, due to the vigorous development of artificial intelligence, many new artificial intelligence algorithms have been applied to problems related to power load forecasting (Friedrich et al.,2014, Koprinska et al., 2015, Yu et al., 2005), especially the emergence of deep learning in the late 2000s. People's interest in neural networks, artificial neural networks, is a structure that imitates biological neural networks (simulating the nervous system of animals), using mathematical models for simulation, and is used to estimate or approximate functions. Neural networks perform calculations by connecting a large number of artificial neurons. In most cases, artificial neural networks can change the weights of internal neurons through changes in external information, and are an adaptive system. It is a nonlinear mathematical prediction model that can learn load sequences and performs well in predicting loads (Fan et al., 2005, Hippert et al.,2001, Hsu & Chen, 2003).

In traditional feedforward neural networks (FNN), the calculated output of each layer will only be passed forward in one direction to the input of the next layer, which means that the input and output are independent and unrelated. One of the more advanced neural networks is recurrent neural network (RNN). The difference between RNN and FNN is that RNN can pass the calculated output of a certain layer back to the layer itself as an input value. In RNN, the trigger value from each time point will be stored in the internal state of the network, in order to provide time storage properties (Bayer & Simon , 2015). However, the main weakness of RNN is that it cannot learn time series related to long-term memory(Bayer & Simon , 2015, Pascanu et al., 2013). RNN can only remember short-term memory, but not long-term memory. RNN will encounter problems when learning time series with long-term dependencies. To overcome the problem of gradient vanishing or exploding, Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997) developed the long-short term memory (LSTM) algorithm as an extension of RNN (Pascanu et al., 2013, Sutskever, 2012).

In order to improve prediction performance, in addition to single prediction algorithms, hybrid models are often used, and models using empirical mode decomposition (EMD) have received great attention (Chen et al.,2012). EMD(Huang et al., 1998) is a useful method for processing nonlinear signal analysis (such as stock data) or other related fields(Vincent et al.,1999, Yu et al., 2018), and provides a new method for processing nonlinear and non-stationary signals. EMD-based forecasting methods have been used in wind speed forecasting (An et al., 2012), industry (Feng et al.,2010), tourism management (Lai & Yeh, 2013) and financial time series forecasting (Fu, 2010). Based on EMD, any complex signal can be decomposed into limited inherent modes. However one of the main disadvantages of EMD is the frequent occurrence of mode mixing, which is defined as a single (IMF), which consists of signals of widely different scales or similar scale signals appear in different IMFs. Modal aliasing is caused by signal interruptions. Interruptions are an indefinite form of disturbance signals, which we often encounter. Interruptions will cause confusion in the time-frequency distribution, thereby destroying the physical meaning of IMF. In response to the mode mixing problem, Huang proposed the ensemble empirical mode decomposition (EEMD) (Wu & Huang, 2009) in 2009. As an improvement, EEMD is a noise-assisted data analysis method. Firstly, EEMD adds white noise to the signal, then performs empirical mode decomposition on the signal, and repeats the above two steps several times to obtain several sets of IMFs. Finally, the respective the intrinsic mode functions are averaged to offset the effects of noise.

In this study, there are three research motivations: First, most of the past prediction models only use a single attribute as the input variable of the power prediction model.

Previous research shows that adding prediction attributes related to prediction in the prediction model can improve the performance of the power prediction model. Second, most traditional time series models use the latest period data with noise as the input variables of the prediction model. However, noise caused by environmental and weather conditions is included in the original input data. In order to overcome the above shortcomings, this study reputes that EEMD can decompose complex original data into simpler frequency components and highly correlated input variables and it can overcome the problem of mode mixing that cannot be handled by the EMD method. Third, the LSTM method can overcome the problem of gradient disappearance or explosion when the RNN algorithm learns time series with long-term dependencies. In past research, the LSTM has been used to deal with the prediction problem of time series and excellent prediction results can be obtained.

Based on above reasons, this study will use autoregression (AR) combined with EEMD method. Proposed model uses EEMD to decompose complex time series raw data into simpler frequency components and highly correlated input variables, and uses the moving average method to generate prediction model input attributes. Proposed model would combine the prediction input attributes generated by AR-EEMD and moving average methods with the LSTM neural network to establish a prediction model for power demand forecasting. The proposed method is based on AR-EEMD and moving average. The long-short-term memory neural hybrid model is applied to power forecasting. The procedure of method is described as follows

(1) Data collection: First, Taiwan's daily domestic electricity load and daily industrial electricity load are collected from government's open data platform. First 10 months (January-October) datasets of each year are used as the training set, and the last two months (November-December) datasets are used as the test set.
(2) Establish AR model: Use the least squares method to detect the number of lagging periods to establish an AR model.
(3) Decompose input variables: Use EEMD to decompose the input variables of the AR model into several IMFs
(4) Establish a moving average model: Calculate the moving average of the electricity time series to establish a moving average model
(5) Establish a prediction model: Use the IMF attribute set generated by AR-EEMD and the value calculated by the moving average method as prediction input attributes and combine it with the LSTM neural algorithm to establish an electricity load prediction model.
(6) Evaluation and comparison: Evaluate the prediction performance and compare the prediction capability with other prediction methods.

Related works

Autoregressive Model

In time series forecasting, forecasts are actually obtained by predicting values in the next time period based on a specific forecasting algorithm. Additionally, forecasting non-periodic short-term time series is much more difficult than forecasting long-term time series. The Autoregressive Moving Average (ARMA) is a traditional method that is well suited for forecasting periodic cyclical data, such as seasonal or cyclical time series (Chang, 2008).

Box and Jenkins (1976) generate a generalized linear stochastic model by assuming that time series data can be linearly aggregated through stochastic fluctuations. In this study, we focus on the AR model, which is a model that contains the past values of one or more explanatory variables of the dependent variable. The simplest AR(1) is defined as

$$y_t = \phi_1 y_{t-1} \quad (1)$$

When considering random errors and constant terms, the improved AR(1) model becomes

$$y_t = \mu + \phi_1 y_{t-1} + u_t \quad (2)$$

Where ϕ_1 is the first-order autoregressive coefficient and u_t is white noise, which is regarded as a random error. An autoregressive model is simply a linear regression of the current value of a series against one or more previous values of the series. In the AR(1) model, it can be considered that for a given value y in time period *t*, there is a relationship with time period *t-1*. If there is an autoregressive model of order p, the AR(p) model can be expressed as

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + u_t \quad (3)$$

Ensemble Empirical Mode Decomposition

Huang et al. proposed the EEMD signal decomposition algorithm in 2009 (Wu & Huang, 2009), which has been widely used in various fields in recent years(Tan et al., 2018, Wang et al., 2019, Liu et al., 2019, Chen & Wang, 2018, Yang et al., 2016). The main purpose of EEMD is to decompose the original complex time series from Remove the IMF. Mathematically, the IMF must satisfy the following two conditions: (1) The sum of the numbers of local maxima and local minima must be equal to the number of zero crossings or It can only differ by 1 at most, which means that an extreme value must be immediately followed by a zero-crossing point. (2) At any point in time, the average of the upper envelope defined by the local maximum and the lower envelope defined by the local minimum is close to zero.

In order to avoid modal aliasing, EEMD performs EMD multiple times on the original time series x(t), (t = 1, 2, ..., T), given some Gaussian white noise to obtain a set of IMFs, and then The overall average of the corresponding IMF is regarded as the final decomposition result. The main steps of EEMD are as follows

(a) Add the white noise sequence $w^i(t) \sim N(0, \sigma^2)$ to the original time series $x(t)$ to construct a new series, as follows

$$x^i(t) = x(t) + w^i(t) \quad (4)$$

(b) Use EMD to decompose the time series $x^i(t)$ with white noise added into n IMFs $c_j^i(t)(j = 1,2,\dots,n)$ and a residual $r^i(t)$. The formula is as follows

$$x^i(t) = \sum_{j=1}^{n} C_j^i(t) + r^i(t) \quad (5)$$

where $c_j^i(t)$ is the *j-th* IMF in the *i*-th test

(c) Repeat step (a) and (b) each time to give a certain number M of different times of white noise to obtain the corresponding IMFs.

(d) Calculate the IMFs corresponding to the average of M tests as the final IMFs

$$c_j^i(t) = \frac{1}{M}\sum_{i=1}^{M} c_j^i(t) \tag{6}$$

Once EEMD is complete, the original time series can be represented as a linear combination of IMFs and residuals as follows

$$x(t)=\sum_{j=1}^{n} c_j(t) + r(t) \tag{7}$$

where $c_j(t)$, $(t = 1,2,...,T)$ is the j-th IMFs extracted during the j-th decomposition process at time t, $r(t)$ is the final residual, and n is the number of IMFs.
Assuming that complex data in reality involves real information and noise, and that the integrated average of data with different noise is closer to the real signal, white noise time series are used to capture the real IMF and offset themselves via the ensemble average. The blocking of final errors can be controlled by adding white noise

$$\varepsilon_{ne} = \frac{\varepsilon}{\sqrt{M}} \tag{8}$$

where M is the number of integration members, ε is the amplitude of the additional noise sequence, and ε_{ne} is the final standard error deviation, defined as the difference between the input signal and the corresponding IMF.

LSTM, long-short term memory

Long-short term Memory (LSTM) is a temporal recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber(1997) in 1997. Due to the unique design structure, LSTM is suitable for processing and predicting important events with very long intervals and delays in time series. In recent years, it has been used in deep learning, artificial intelligence and various applications in different fields (Zhu et al., 2018, Wang et al., 2018, Yu et al., 2018, Zhao et al.,2019, Tian et al., 2018), the traditional recurrent neural network (RNN, Recurrent Neural Network) is one of the recursive neural network methods that can be used for continuous data modeling. The key feature of RNN is the network delay recursion, which enables it to describe the dynamic performance of the system (Bayer & Simon, 2015), the signal delay recursion makes the output of the network at time t related not only to the input at time *t*, but also to the recursive signal before time *t*.

Although RNN can handle short-term sequence data, when RNN is used to deal with problems that need to consider long-term data dependencies, the prediction ability of RNN is poor(Pascanu et al., 2013, Sutskever, 2012). In the past, scholars have proposed several different RNNs. An improved algorithm for the model, but the LSTM method uses dedicated LSTM storage units to represent long-term dependencies in time series data, making LSTM the most commonly used prediction algorithm to replace the RNN method. In addition, the LSTM method can solve the vanishing gradient problem of RNN(Pascanu et al., 2013, Hochreiter & Schmidhuber, 1997). LSTM is a model proposed by Horchreiter and Schmidhuber(1997). This method uses a special multiplication unit to control the constant error and force the constant error information flow direction to truncate the network gradient. These nonlinear units learn to open or close gates in the network in order to regulate this constant flow of errors (Sutskever, 2012). Therefore, LSTM adopts the method of properly considering long-term memory delay information to improve the traditional RNN algorithm(Tang et al., 2017).

The key to the LSTM structure is the unit state (storage unit), which looks like a conveyor belt. It runs directly along the entire chain, capable of adding or removing information to the unit state, carefully regulated by structures called valves. A valve is a pipe that serves as a selective entry point for information. They consist of sigmoid neural network layers and pointwise multiplication operations. The input value at time point t is (X_t), and the previous time step (S_{t-1}) is introduced into the block of the LSTM. The hidden layer (S_t) is calculated as follows:

The first step in LSTM is to determine what information will be discarded from the memory cell state. This decision is made by the following forget valve $(f\,t)$:

$$f\,t = \sigma(X_t U^f + S_{t-1} W^f + b_f) \quad (9)$$

The following steps are to decide what new information to store in the cell state. This step has two loops: first, the input valve (i_t) layer decides which values to update. Second, the tanh network layer creates a vector of new candidate values $\widetilde{C_t}$. These two loops can be described as follows:

$$i_t = \sigma(X_t U^i + S_{t-1} W^i + b_i) \quad (10)$$

$$\widetilde{C_t} = \tanh(X_t U^c + S_{t-1} W^c + b_c) \quad (11)$$

Then, the old memory cell state C_{t-1} is updated to the new cell state C_t, and its calculation formula is as follows:

$$C_t = C_{t-1} \otimes f\,t \ \oplus\_ i_t \otimes \widetilde{C_t} \quad (12)$$

In the end, it will be decided which information will be output. This output result is based on the state of the memory unit, but the result is the message generated after filtering. In this step, the output valve (o_t) determines which parts of the memory cell state will be treated as output values. The memory cell state is then passed through the tanh layer (converting the value to between -1 and 1) and its value is multiplied by the output valve, the formula is as follows:

$$o_t = \sigma(X_t U^o + S_{t-1} W^o + b_o) \quad (13)$$

$$S_t = o_t \otimes \tanh(C_t) \quad (14)$$

From the previous six equations, LSTM provides the following three sets of parameters:

1. Input weights: $\boldsymbol{U^f}$, $\boldsymbol{U^i}$, $\boldsymbol{U^o}$, $\boldsymbol{U^c}$.
2. Recursive weights: $\boldsymbol{W^f}$, $\boldsymbol{W^i}$, $\boldsymbol{W^o}$, $\boldsymbol{W^c}$.
3. Bias: $\boldsymbol{b_f}$, $\boldsymbol{b_i}$, $\boldsymbol{b_o}$, $\boldsymbol{b_c}$.

Proposed Model

This study mainly constructs a power prediction model based on autoregressive, EEMD model and moving average method, combined with LSTM neural network. The developed system and results can be applied in government energy policy planning and industrial electricity demand assessment. Forecasting the peak moments of people's electricity consumption can prevent blackouts and avoid energy waste caused by over-production of electricity. Taiwan's electricity generation still relies on coal-fired and thermal power generation. Waste of electricity production not only causes a waste of resources, but also causes environmental pollution. As well as the increase of PM2.5 and the generation of smog, it will have a great impact on the economy, environmental protection and people health. The research structure of this plan is shown in Figure 1.

In the past discussion of literature related to power forecasting, three main shortcomings were found: (1) In the past, power forecasting models mostly used the most recent data as the input variables of the forecasting model. However, the raw data contained noise, which would reduce the accuracy of the forecast. In the past, Research has used the EMD method to decompose complex time series raw data into simpler frequency components and highly correlated input variables. However, the EMD method cannot handle the mode mixing problem; (2) Past predictions Most models only use a single attribute as the input variable of the power prediction model. Previous research shows that adding prediction attributes related to prediction in the prediction model can improve the prediction accuracy; and (3) the RNN algorithm is learning long-term dependencies. When handling with time series data, RNN will encounter the problem of vanishing or exploding gradients. In order to solve these shortcomings, this study uses the autoregressive model to analyze the lag periods of the time series and uses EEMD to decompose the input variables (EEMD can solve the problem of modal aliasing), and uses the moving average model to generate the numerical change trend of the series. Proposed model finally combined generated features with LSTM neural network (which can solve the problems caused by RNN when learning time series of long-term dependencies) to establish a new prediction method.

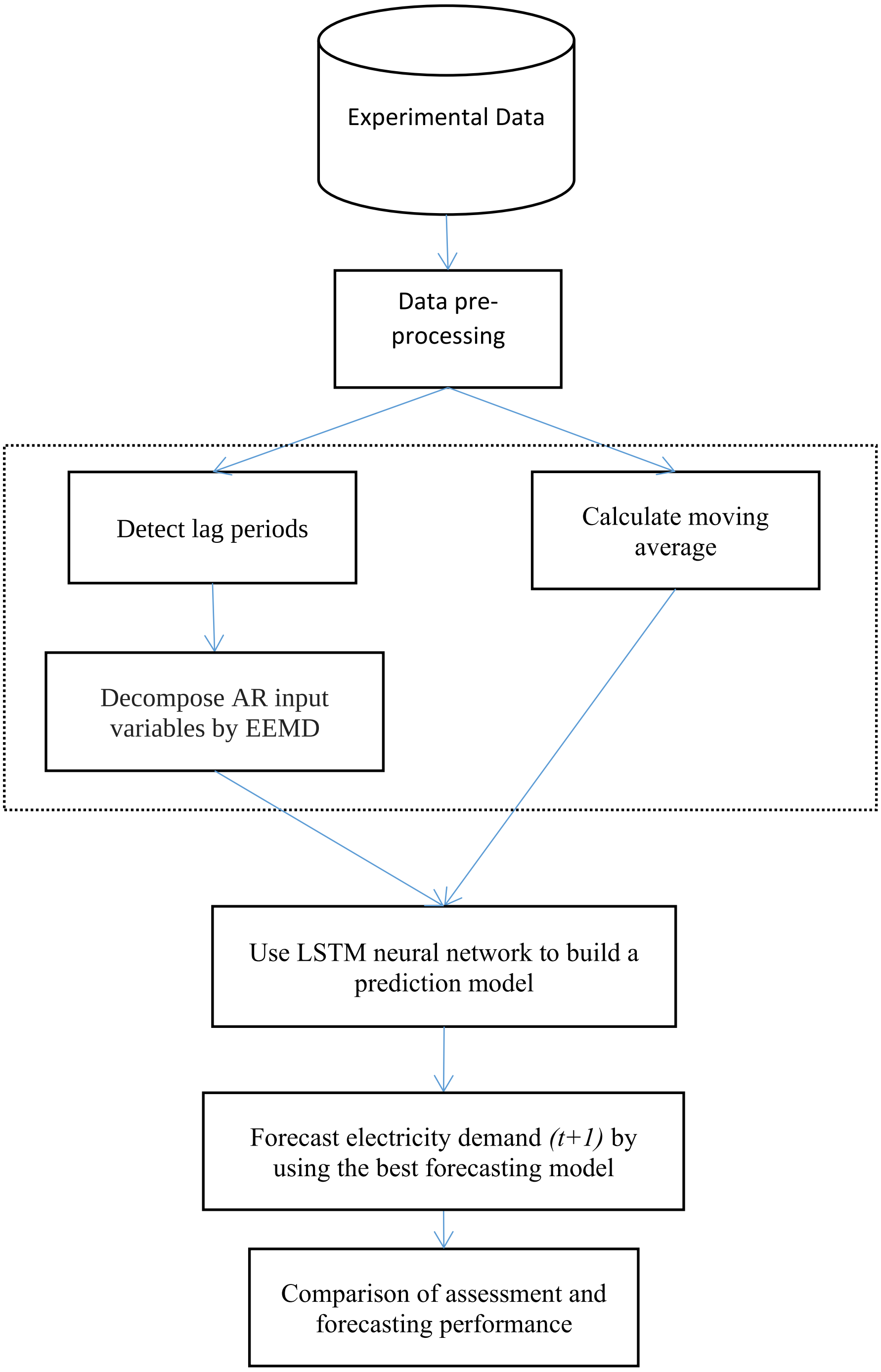

Figure 1. Architecture of proposed prediction model

The structure of this research is shown in Figure 1. First, the data set is collected and then implements data pre-processed procedure. The lag period verification is performed to establish the AR model. EEMD is used to decompose the lag period data of the AR model, and then the moving average is calculated. Then the IMFs generated by AR-EEMD and the values calculated by the moving average method are used as prediction variables, and the LSTM neural network is further used to establish a power demand prediction model. The training data set is used to generate the best prediction model, and then the generated Forecasting models perform power demand forecasting. The calculation steps are as follows:

Step 1: Collect Data Set
Daily industrial electricity consumption data and domestic electricity consumption data will be collected as experimental data. The data for the first 10 months of each year (January to October) will be used as the training set, and the last two months of each year (November to December) is used as a test data set, and the data source is collected from the government data open platform (https://data.gov.tw/).

Step 2: Data pre-processing
In general, data may have inconsistencies, errors, out-of-range values, impossible data combinations, missing values or noise. The above data are not suitable for the data mining process. Therefore, this step removes all missing values and then reformats the dataset into the format of the data mining algorithm.

Step 3: Detect lag periods - establish AR mode
In this step, the study will use E-Views software to characterize the AR model for different lag periods and levels of electricity loads (EL). Past experimental results show that power demand forecasting is highly related to recent power demand data. Therefore, this study uses the linear regression variable of the power load value of the last five periods of forecast data from EL *(t-1)* to EL (*t-5*) for estimation and testing. Using the *p* value to perform hypothesis testing to establish the number of levels of the AR model.

Step 4: Decompose AR input variables by EEMD
In the third step, this study established the AR model by verifying the number of lag periods of the electric load. Then, EEMD was used to decompose the input variables of the AR model into a set of finite numbers with simpler frequency components and stronger correlations.

Step 5: Calculate moving average (*MA*)-establish a moving average model
The moving average is used to help identify power demand trends. Using the moving average as a forecast attribute will improve the forecasting performance of power forecasting. The predicted value of power demand time forecast is closely related to the recent power demand value, therefore this study calculates the moving average of three periods and five periods as the input attribute of the prediction model.

The formula of *MA* is as follows

$$MA = \frac{SUM_{(N)}(\mathrm{EL})}{DAY(N)} \tag{15}$$

EL: Daily electrical load;
DAY(N): The number of previous periods *(N)*.
$SUM_{(N)}$: The sum of the electric load of the previous periods, period *(N)*

Step 6: Use LSTM neural network to build prediction model
This study uses the values generated by AR-EEMD and the moving average method as input variables of the LSTM neural network to establish electricity demand prediction model, and utilizes RMSE (root-mean-square error) such as Equation (16) as evaluation indicator.

Proposed model uses the training data to select the best prediction model based on the situation that produces the minimum RMSE.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} |A(t) - F(t)|^2}{n}} \tag{16}$$

A(t) is the actual value at time *t*, *F(t)* is the predicted value at time *t*, *n* is the number of data

Step 7: Use the best prediction model to predict electricity demand *(t + 1)*
In step 5, using the training data and under the condition of minimum RMSE, we can obtain the best parameter values, and then proposed model uses the prediction model with the best parameter values to predict test dataset.

Step 8: Comparison of evaluation and prediction performance
In this step, proposed model uses equation (16) to calculate RMSE under test dataset, and utilizes RMSE of all test data sets as evaluation indicator to compare with other prediction methods

Experiments and Comparisons

In this section, daily industrial electricity consumption (unit: million kilowatt hours) and domestic electricity consumption (unit: million kilowatt hour) collected in 2016, 2017, and 2019) is used as an experimental dataset to verify the prediction accuracy of the proposed method. The data in the first 10 months of each year (January to October) will be used as the training set, and the last two months of each year (November to December) as test data set.

This project will use E-Views software to characterize the AR model for different lag periods and levels of electricity loads (EL). Therefore, this study uses power load value in the last five periods of forecast data from EL (t-1) to EL (t-5), for estimation and testing. Proposed model utilizes daily industrial electricity consumption in 2017 year as test data set and uses the *p* value to do hypothesis testing to establish the number of levels of the AR model. The results of the E-Views software verification are shown in Figure 2. The figure shows that the p-values from EL (t-1) to EL (t-5) are all less than the significant level of 0.05, so the number of levels of the AR model is 5th order.

Dependent Variable: ELECTRICITY_LOAD
Method: Least Squares
Date: 08/20/20 Time: 15:12
Sample (adjusted): 6 365
Included observations: 360 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884826	5.074432	1.947967	0.0522
ELECTRICITY_LOAD(-1)	1.072381	0.051131	20.97314	0.0000
ELECTRICITY_LOAD(-2)	-0.513109	0.075697	-6.778451	0.0000
ELECTRICITY_LOAD(-3)	0.356574	0.078409	4.547602	0.0000
ELECTRICITY_LOAD(-4)	-0.224370	0.075735	-2.962588	0.0033
ELECTRICITY_LOAD(-5)	0.270860	0.051009	5.310013	0.0000

Figure 2. Calibration of daily industrial electricity consumption lag times

Then proposed uses EEMD to decompose the AR input variables. Because the order of the AR model is 5th order, in order to avoid decomposing too many IMFs, the maximum number of decompositions of the input variables of each AR model is set to 2, plus the decomposition will Residuals are generated (residuals are also treated as 1 IMF), so a total of 15 IMF values are decomposed and used as input variables of the prediction model, and the moving average of the first three periods and the first five periods is calculated as the input of the prediction model. This study uses keras deep learning library of the python program to establish an LSTM prediction model. The model summary is shown in Figure 3. It includes an LSTM hidden layer containing 128 neural units, and a dense neural network layer (containing 128 neural units). The activation function is relu (rectified linear unit), and there is an output layer. Proposed model set MSE (mean squared error) as the loss function. The optimization method during training is adam, and evaluation method is mae (mean absolute error). Each batch is 128 pieces of data, and the training cycle is 400 times. The method proposed in this study is the AR-EEDM-MV-LSTM prediction model, and forecasting performance is compared with other models. Proposed model uses the AR model of the first five periods and uses EEMD for attribute decomposition, and then calculates the moving average of periods 3 and 5. This study uses RNN to build a prediction model (including an RNN hidden layer containing 128 neural-like units, a Dense neural network layer (containing 128 neural-like units, the activation function is relu (rectified linear unit)). Proposed model sets mse (mean squared error mean square error) as the loss function, the optimization method during training is adam. Evaluation model method is mae (mean absolute error), each batch is 128 data, the training cycle is 400 times). Another comparison model is AR-MV-RNN (using the AR model of the first five periods, then calculating the moving average of periods 3 and 5, and using RNN to build a prediction model), the setting method of the RNN model is the same as The RNN model in AR-EEDM-MV-RNN is the same, The other comparison model is AR-MV-LSTM (using the AR model of the first five periods, then calculating the moving average of periods 3 and 5, and using LSTM to build a prediction model). The design of the LSTM model is the same as the LSTM model in AR-EEDM-MV-LSTM.

```
_______________________________________________________________
__
Layer (type)                 Output Shape              Param #
===============================================================
==
lstm_17 (LSTM)               (None, 128)               66560
_______________________________________________________________
__
dense_64 (Dense)             (None, 128)               16512
_______________________________________________________________
__
dense_65 (Dense)             (None, 1)                 129
===============================================================
==
Total params: 83,201
Trainable params: 83,201
Non-trainable params: 0
```

Figure 3. Summary diagram of LSTM prediction model

Next, this project compares two data sets of industrial electricity consumption and civilian electricity consumption: the performance of the proposed method is compared with other methods.

Industrial Electricity Dataset

This section uses three years of industrial electricity consumption data in 2016, 2017, and 2019 as our test data set, and combines it with the other three prediction methods (AR-MV-RNN, AR-MV-LSTM, AR-EEDM-MV- RNN), and MAE (experimental results are shown in Table 1) and RMSE (experimental results are shown in Table 2) are used as evaluation formulas respectively. From the experimental results in Table 1 and Table 2, it can be seen that the method obtained in this study is better than to other forecasting methods.

Table

Experimental results of different prediction models for industrial electricity consumption data sets (based on MAE evaluation index)

Year / Model	2016	2017	2019
AR-MV-RNN	15.781	16.987	13.708
AR-MV-LSTM	14.596	14.915	11.750
AR-EEDM-MV-RNN	15.339	16.906	12.916
Propose model (AR-EEDM-MV-LSTM)	13.520*	13.996*	11.691*

*Experimental results with the best performance among the four prediction models

Table 2

Experimental results of different prediction models for industrial electricity data sets (based on RMSE evaluation index)

Year / Model	2016	2017	2019
AR-MV-RNN	17.844	20.552	17.667
AR-MV-LSTM	16.661	18.192	15.739
AR-EEDM-MV-RNN	17.623	20.537	17.057
Propose model (AR-EEDM-MV-LSTM)	15.685*	17.783*	15.729*

*Experimental results with the best performance among the four prediction models

Civilian electricity consumption data set

This section uses the three-year civilian electricity consumption data in 2016, 2017, and 2019 as our test data set. The other three prediction methods (AR-MV-RNN, AR-MV-LSTM, AR-EEDM-MV- RNN) are as comparison models, and MAE (experimental results are shown in Table 3) and RMSE (experimental results are shown in Table 4) are used as evaluation indictors respectively. From the experimental results in Tables 3 and 4, it is known that the method obtained in this project is better than to other forecasting methods.

Table 3
Experimental results of different prediction models for civilian electricity consumption data set (based on MAE evaluation index)

Year / Model	2016	2017	2019
AR-MV-RNN	26.190	30.967	32.706
AR-MV-LSTM	25.584	29.642	31.541
AR-EEDM-MV-RNN	24.831	30.945	32.423
Propose model (AR-EEDM-MV-LSTM)	21.323*	29.338*	29.417*

*Experimental results with the best performance among the four prediction models

Table 4
Experimental results of different prediction models for civilian electricity consumption data set (based on RMSE evaluation index)

Year / Model	2016	2017	2019
AR-MV-RNN	29.182	33.651	38.776
AR-MV-LSTM	28.365	32.169	37.791
AR-EEDM-MV-RNN	27.776	33.144	38.232
Propose model (AR-EEDM-MV-LSTM)	25.862*	32.123*	37.641*

*Experimental results with the best performance among the four prediction models

Conclusions

The hybrid method proposed in this study uses EEMD and LSTM combined with deep learning technology to predict industrial and domestic electricity demand. Experimental results show that proposed model can improve the accuracy of prediction. From Table 1 to Table 4, the prediction performances of the proposed model are better than that of AR-MV-LSTM without EEMD. AR-EEDM-MV-RNN (using EEMD method) has better prediction performance than the AR-MV-RNN method without EEMD. It is obvious that the EEMD method can decompose the original data containing noise into simpler components and highly correlated input variables, which can be effective reduced forecast errors. In addition, it can be seen from Tables 1 to 4 that prediction performance of proposed model is better than the method AR-EEDM-MV-RNN without LSTM, and the AR-MV-LSTM method using LSTM is better than the method AR-MV-RNN (using RNN) has better prediction performance. From the experimental results, it is known that the LSTM method can solve the problem of gradient disappearance or explosion when the RNN algorithm learns time series of long-term dependencies and improves the prediction accuracy.

Major finding are summarized as follows

1. Experimental results indicate that EEMD method can could reduce forecasting error more beneficially than EMD method model.
2. Experimental results show that LSTM method can could improve prediction accuracy than EMD method model.

Suggestions of this paper are as follows: The forecasting model proposed in this study can be applied to other related fields such as product demand forecasting, unemployment rate forecasting, and consumer price index forecasting. For following research, researchers can use other electricity loads datasets to further validate the proposed model.

References

An, X., Jiang, D., Zhao, M., Liu, C. (2012) Short-term prediction of wind power using EMD and chaotic theory, Commun. Nonlinear Sci. Numer. Simul. 17 1036-1042.

Bayer, F. M., Bayer, D. M., Pumi, G. (2017) Kumaraswamy autoregressive moving average models for double bounded environmenttal data, Journal of Hydrology, 555, 385-396,

Bayer, J., Simon. (2015) Learning Sequence Representations, Technische Universitt Mnchen.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics. 31 307-327.

Box, G., Jenkins, G. (1976) Time series analysis: Forecasting and control, San Francisco: Holden-Day.

Caruana, R., A. Niculescu-Mizil, Crew, G., Ksikes, A., (2004) Ensemble selection from libraries of models, International conference on machine learning,

Chang, B. R. (2008) Resolving the forecasting problems of overshoot and volatility clustering using ANFIS coupling nonlinear heteroscedasticity with quantum tuning. Fuzzy Sets and Systems, 159(23) 3183-3200.

Chen, J., Wang, Y. L. (2018) A Resource Demand Prediction Method Based on EEMD in Cloud Computing, Procedia Computer Science, 131, 116-123

Chen, K. L., Yeh, C. C., Lu, T. (2012) Forecasting the output of Taiwan's integrated circuit (IC) industry using empirical mode decomposition and support vector machines, Int. J. Phys. Sci. 3 (78) 5460-5467.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimator of the variance of United Kingdom inflation. Econometrica. 50(4) 987-1008.

Fan, S., Mao, C., Chen, L. (2005) Peak load forecasting using the selforganizing map. "*Advances in Neural Network*"-ISNN 2005. Springer. Part III. 640-9

Feng, F., Zhu, D., Jiang, P., Jiang, H. (2010) GA-EMD-SVR condition prediction for a certain diesel engine, 2010 Progn. Syst. Heal. Manag. Conf. PHM '10,

Friedrich, L., Armstrong, P., Afshari, A. (2014)Mid-term forecasting of urban electricity load to isolate air-conditioning impact, "*Energy and Buildings*", 80, September, 72-80,

Fu, C. (2010). Forecasting exchange rate with EMD-based Support Vector Regression, in: 2010 Int. Conf. Manag. Serv. Sci. MASS,

Hippert, H. S., Pedreira, C. E., Castro, R. (2001) Neural networks for short-term load forecasting: a review and evaluation. *"IEEE Trans Power Syst",* 16, 44-55,

Hochreiter, S., Schmidhuber, J. (1997) Long short-term memory, Neural Comput. 9 (8) 1735–1780.

Hsu, C. C., Chen, C. Y. (2003) Regional load forecasting in Taiwan- applications of artificial neural networks. *"Energ Convers Manage"*, 44: 1941-9

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q. (1998)The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, in: Proceedings of the royal society of London series a-mathematical physical and engineering sciences, series A, 454 903-995.

Huarng, K. H. (2001) Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems. 123 155-162.

Koprinska, I., Rana, M., Agelidis, V. G. (2015) Correlation and instance based feature selection for electricity load forecasting, "*Knowledge-Based Systems*", 82, 29-40,

Lai, M. C., Yeh, C. C. (2013) A hybrid model by empirical mode decomposition and support vector regression for tourist arrivals forecasting, J. Test. Eval. 41.

Song, Q., Chissom, B. S. (1993)Forecasting enrollments with fuzzy time series Part I, Fuzzy Sets and Systems. 54 1-10.

Sutskever, I. (2012) Training recurrent neural networks, University of Toronto, Ph.d. thesis .

Tang, L., Lv, H., Yu, L. (2017) An EEMD-based multi-scale fuzzy entropy approach for complexity analysis in clean energy markets, Appl. Soft Comput. 56 124-133.

Tian, Y., Zhang, K., Li, J., Lin, X., Yang, B. (2018) LSTM-based traffic flow prediction with missing data, Neurocomputing, 318 (27), 297-305

Vincent, H. T., Hu S.-L. J., Hou, Z. (1999) Damage detection using empirical mode decomposition method and a comparison with wavelet analysis, in: Proceedings of the second international workshop on structural health monitoring, Stanford 891-900.

Wang, W., Chen, Q., Yan, D., Geng, D. (2019) A novel comprehensive evaluation method of the draft tube pressure pulsation of Francis turbine based on EEMD and information entropy, Mechanical Systems and Signal Processing, 116(1), 772-786

Wang, J., Peng, B., Zhang, X. (2018) Using a stacked residual LSTM model for sentiment intensity prediction, Neurocomputing, 322 (17), 93-101

Wei, L. Y. (2013) A GA-weighted ANFIS model based on multiple stock market volatility causality for TAIEX forecasting , Applied Soft Computing, 13 911-920

Wei, L.Y. (2016) A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting, Applied Soft Computing, 368-376

Wu, Z., Huang, N. E. (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adapt. Data Anal. 1 (01) 1–41.

Xu, J., Tan, X., He, G., Liu, Y. (2019) Disentangling the drivers of carbon prices in China's ETS pilots - An EEMD approach Technological Forecasting and Social Change, 139 , 1-9

Yang, Y., Chen, Y., Wang, Y., Li, C., Li, L. (2016) Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting, "*Applied Soft Computing*", Vol. 49, pp 663-675,

Yu, D. J., Cheng, J. S., Yang, Y. (2005) Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Signal Process. 19 (2) 259-270.

Yu, Z., Liu, G., Liu, Q., Deng, J. (2018)Spatio-temporal convolutional features with nested LSTM for facial expression recognition, Neurocomputing, 317 (23), 50-57

Zhang, C., Z. Ni, Ni, L., J. Li, Zhou, L. (2016)Asymmetric multifractal detrending moving average analysis in time series of PM2.5 concentration, Physica A: Statistical Mechanics and its Applications, 457, (1), 322-330,

Zhao, X. J., Mao, Chen, L. (2019)Speech emotion recognition using deep 1D & 2D CNN LSTM networks, Biomedical Signal Processing and Control, 47, 312-323

Zhu, X., Li, Lixiang, Liu, Jing., Li, Ziyi, Niu, X. (2018)Image captioning with triple-attention and stack parallel LSTM, Neurocomputing, 319 (30), 55-65