A Multifactorial Framework for Short-Term Load Forecasting System as Well as the Jinan's Case Study

YA GAO, YONG FANG, HUANHE DONG, AND YUAN KONG
College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China
Corresponding author: Yuan Kong (kongyuan@sdust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11975143, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2019QD018.

ABSTRACT Accurate and reliable short-term electric load forecasting (STLF) plays a critical role in power system to enhance its routine management efficiency and reduce operational costs. However, most of the existing STLF methods suffer from lack of appropriate feature selection procedure. In this paper, a multifactorial framework (MF) possessing the potential to contribute more satisfactory forecasting results and computational speed is proposed. Moreover, a graphical tool for easy and accurate computation of day-ahead load forecast is implemented via MATLAB App Designer. Firstly, we choose the candidate feature set by analyzing the raw electricity consumption data. Then, partial mutual information is adopted as criterion to eliminate these irrelevant and redundant ones among candidate features for the purpose of reducing the input subset and retaining these most relevant. At last, the selected features are used as the input of the well-established artificial neural network (ANN) model optimized by genetic algorithm and cross validation to implement prediction. The MF is applied for the load data measured from 2016 to 2018 in Jinan, and then some competitive experiments and extensive simulations are carried out and results indicates that the ANN-based model with selected features significantly outperforms other alternative models with single features or a few of features regarding mean absolute percent error. In addition, the parallel structure of ANN and the lower dimension of the input space enable the model to achieve faster calculation speed.

INDEX TERMS Short-term load forecasting, feature selection, partial mutual information, artificial neural network.

I. INTRODUCTION
Short-term electric load forecasting (STLF) is an important issue for the planning and management operation of power system and as the basis of energy transaction and decision in the competitive energy market. The accuracy of forecasting result is a very crucial factor to make most predictions for future demands from energy sector [1], [2]. The generator can be run at the lowest cost when the load demand is known in advance. As said in [3], a small increase in load forecasting accuracy will save the company millions of dollars. However, load demand is also a non-linear and non-stationary process affected by various factors, which complicates forecasting work [4]. First, the load series is highly complex and exhibits several levels of seasonality: the load at a given hour is dependent not only on the load at the previous hour, but also on the load at the same hour on the previous day, and on the load at the same hour on same day in the previous week. Secondly, there are many important exogenous variables that must be considered, specially weather-related variables, such as temperature and humidity.

Many well-known approaches have been proposed for STLF to continue improving forecasting performance in past decades. The existing STLF methods can be of three types, the first one is based on the statistical methodology, another involves artificial intelligence technology, and the rest is the hybrid model.

A. STATISTICAL FORECASTING MODELS
Statistical methods often use historical data to look for the correlation between exogenous factors mentioned above and electric load. In the early stages of STLF, statistical
methods were extensively employed, such as regression models [5], [6], exponential smoothing [7], autoregressive moving average model (ARMA) [8] and autoregressive Integrated moving average model (ARIMA) [9]. These statistical approaches with lower calculation are relatively easy to be established and implemented. However, these approaches are difficult to achieve substantial improvements owing to their theoretical definitions, which largely limit their forecasting ability and cannot receive the expected forecasting accuracy [10].

B. ARTIFICIAL INTELLIGENCE FORECASTING MODELS

Due to the superior nonlinear computing capability, artificial intelligence (AI) techniques (e.g. artificial neural networks (ANN) [11]–[13], fuzzy logic models [14], and support vector machines (SVM) [1], [15]) have been applied to cope with the STLF problem. The most representative technique is ANN, which is suitable for STLF because of its ability of nonlinear mapping and generalization. In [13], by comparing several forecasting methods, including both large neural networks and conventional regression-based methods, it is found that good performances for the large neural networks are not only with the smallest mean absolute percentage error value (MAPE) (2.35-2.65%), but also with a lesser spreading of the errors. A combination of fuzzy time series with seasonal autoregressive fractionally integrated moving average is proposed in [14]. The analysis of the results indicate that the proposed approach presents higher accuracy than any other counterpart. Chen et al. [15] proposed a new support vector regression (SVR) based STLF approach with the ambient temperature of two hours as input variables and electric loads from four typical office buildings in China. The simulation results confirm that the newly model significantly receives the highest forecasting performance and stability. However, these proposed artificial intelligence models also have some disadvantage. It is often subjectively determined the network structure, and it is easy to fall into the local optimum during the training process.

C. HYBRIDE FORECASTING MODELS

In recent years, various hybrid or combined models have also been developed for improving the forecasting accuracy of STLF [16], such as (1) the hybridization or combination of these AI models with each other [17]; (2) the hybridization or combination of these AI models with statistical models [18]; (3) the hybridization or combination of these AI models with superior evolutionary algorithms [19], [20]. Among them, the third type is widely applied in the field of STLF. For instance, aiming at improving the accuracy and speed of STLF, bacterial colony chemotaxis is introduced to optimize the parameters of least squares support vector machine (LS-SVM) in [19]. The simulation results show that the proposed approach can achieve higher forecasting accuracy and faster speed than ANN and LS-SVM with grid search. In addition, Zhang et al. [20] propose a novel load forecasting framework by hybridizing self-recurrent SVR with the variational mode decomposition and improved cuckoo search algorithm. Two real-world datasets are used to examine that the performance of the proposed forecasting model significantly outperforms other alternative models.

D. FEATURE SELECTION

Most of the existing load forecasting work focuses on the improvement and reasonable combination of the existing models. However, the selection of input features for STLF usually depends on the daily experience and speculation of decision makers [21]. As we all known, the change of power load demand is affected by internal and external factors. on the one hand, the load series is highly complex and exhibits several levels of seasonality. On the one hand, there are many important exogenous variables that must be considered, specially weather-related variables. Feature selection is a key step in building reasonable forecasting model and has been proved its importance in many research literatures [22]–[24]. Therefore, decision makers should not only consider the selection of appropriate prediction model, but also determine important internal and external input variables [2]. Of course, the impact of these two kinds of influence factors on load varies in different areas. For example, whether atmosphere related factors have significant influence on load depends on actual regional and climatic conditions. But we all agree that temperature is the most important weather effect. However, the complexity lies in the fact that the feature algorithm is applied based on the following considerations: simplicity, stability, number of reduced features, classification accuracy, storage and computational requirements.

The existing methods of feature selection mainly include two categories: filter method and wrapper method. The filter method selects feature subsets based on evaluation criteria like mutual information (MI), correlation analysis (CA), principle component analysis, and numerical sensitivity analysis [21], [25]–[28]. As a result of MI measures the arbitrary dependence between random variables, it is suitable for the ‘information content’ evaluation of features in complex classification tasks. Therefore, MI is not just widely used in feature selection of load forecasting, also in various fields [29], [30]. For example, MI was adopted to select a subset of the most relevance and non-redundant inputs among the candidates for proposed neural network forecasting model in [25]. And experiments show that the neural network model based on feature selection is better than any other model. Different from filter methods, wrapper methods select the appropriate feature subset from the candidates based on the forecasting accuracy. So, meta heuristics algorithm such as the BinJaya algorithm [26], simulated rebounding algorithm [28] and simulated annealing [29] have been developed to improve the search ability, especially when there are many candidate features. For instance, a novel BinJaya algorithm with kernelized fuzzy rough sets is proposed in [26] to select an optimal feature subsets from the entire feature space constituted by a group of system-level classification
features extracted from phasor measurement units data. The
method can effectively solve the feature selection problem
of pattern-recognition-based transient stability assessment.
Recently, a hybrid filter-wrapper approach is proposed to
complement wrapper methods and filter methods with their
inherent advantages [27]. Firstly, the filter method is used
to eliminate the irrelevant and redundant features to form
an input subset of dimension reduction. Then the wrapper
method is applied to the dimension reduction subset to obtain
a set of small features with high prediction accuracy. Through
the hybrid method, appropriate feature variables are selected
as the input of SVR model. Results also confirm that the
proposed hybrid filter-wrapper model has better performance
than other existing models. In [36], a prediction model com-
bined with periodic and non-periodic features is proposed
and a case study is conducted in Qingdao. Some regular
features are abstracted via the spectral analysis as crucial
predictor variables, and weather factors are filtered out by
mutual information method as two important weather factors
to improve the prediction accuracy. The comparative results
of five different experimental results demonstrate the model
considering internal characteristics of load data and external
important influences and non-periodic factors outperform the
others with one single or few factors and more suitable for
Qingdao.

**E. CONTRIBUTIONS OF THIS PAPER**

In this paper, we propose a multifactorial framework (MF)
of ANN based on data analysis and the filter method,
which combines feature selection procedure with forecasting
model construction. Firstly, raw electric load series between
2016 and 2018 of Jinan city is analyzed in detail to developed
some candidate features. Then, the partial mutual informa-
tion (PMI) based filter method is applied to eliminate irrele-
vant and redundant features for the purpose of reducing input
subset. After above two steps, the PMI values corresponding
to some selected features are used as the initial weights of
the input nodes of the ANN prediction model. At last, some
comparative experimental study are carried out to confirm
the prediction performance of the ANN-based model with
selected factors in Jinan. And data 2016 to 2017 has been as
the training set and data from 2018 is used to examine the
performance of the model on out-of-sample data.

The leading contributions of this paper are summed up bellow:

1. A MF for STLF is proposed, which simultaneously
consider feature selection and modeling procedure. The
purpose of this MF is to reasonably adjust the predictors
and establish forecasting models according to the actual situation
of the area to be investigated, thereby achieving the satisfied
forecasting accuracy and faster speed.

2. Several detail and organized analysis forms including
spectral analysis, box-plot analysis and so on are adopted for
finding out internal movement law among load series and
external factors influencing electric demand.

3. To overcome the subjectivity when constructing the
structure of neural network, genetic algorithm is applied to
optimize initial weights and thresholds and cross-validation
to determine the number of hidden layers and corresponding
neurons.

4. Five comparative experiments have been designed and
implemented based on the climate, topography and economic
development, the simulation results are analyzed in different
forms to examined the applicability of the MF forecasting
model in Jinan.

5. A graphical tool for easy and accurate computation
of day-ahead system electric load forecast with MATLAB
App Designer is developed.

**F. ORGANIZATION OF THIS PAPER**

The rest of this paper is organized as follows. In Section 2,
we elaborate on the proposed multifactorial framework for
load forecasting. In Section 3, details about the experi-
mental setting such as dataset, candidate features, accuracy
measures, and selected counterparts for performance testing
can be found. And the experimental results is presented in
Section 4. Finally, discussion and conclusions are shown in
Section 5 and Section 6, respectively.

**II. THE PROPOSED MULTIFACTORIAL FRAMEWORK**

This section describes the proposed multifactorial framework,
which is mainly composed of three parts including
raw data analysis, feature selection based the filter method
and take selected features as input of ANN-based forecasting
method optimized by genetic algorithm and cross validation.

**A. THE PMI BASED FILTER METHOD**

Compared with the wrapper method [27] for feature selec-
tion, the filter method has a faster calculation speed and
lower cost. Nonlinear relationship is the common problem in
STLF modeling. The model based on the linear correlation
between two variables almost cannot detect and quantify the
nonlinear relationship well. Sharma proposed an input deter-
mination method based on the PMI to overcome the limitation
of the correlation coefficient in selecting appropriate model
inputs [31]. The PMI criterion is applied to identify the opti-
mal combination of rainfall predictors among selected ENSO
indices. It can be regarded as a model-free method because it
can fully capture the linear or non-linear correlation between
two variables and does not require any major assumptions
about the basic model structure. In fact, PMI criterion is an
extension of mutual information (MI) concept [32]. MI is
a common criterion for measuring the correlation between
variables and has widely used for input feature selection.
However, a major issue of redundancy has raised because
MI does not account for the interdependency among candi-
date variables directly.

In order to overcome the problems mentioned above,
PMI is adopted to identify candidate features in this paper.
PMI value between output variable Y and input variable X,
for a set of pre-existing inputs $Z$, can be given by
\[
PMI(X, Y) = \int \int f_{X', Y'}(x', y') \log \frac{f_{X', Y'}(x', y')}{f_X(x')f_Y(y')} \, dx' \, dy'
\] (1)
where
\[
x' = x - E[x|z], \quad y' = y - E[y|z]
\]
where $E[-]$ denotes the expectation operation. $f_{X'}, f_Y$ and $f_{X', Y'}$ are respective univariate and joint probability densities estimated at the sample data points. The variables $x'$ and $y'$ only contain the residual information after the effect of the pre-existing set of inputs $Z$ has been taken into consideration by using the conditional expectations. In feature selection based on PMI, the input variable with the highest PMI value is added as a new predictors. Detailed process of PMI can be found in [33]. Here, we briefly outline the PMI based input feature selection procedure for our proposed approach:
1) Initialize: Set $X$ to be the candidate inputs, $Z$ to the predictors set of inputs, $Y$ to output;
2) Estimate the PMI scores: Compute the PMI $(X, Y)$ between output variables $Y$ and each of the variable in candidate set $X$;
3) Input select: Identify the input $x$ with the highest PMI in step 2; If this PMI score is higher than the 95th percentile randomized sample PMI score, add $x$ to the predictors set, and remove it from $X$. If it is not significant or there is no input in $X$, go to Step 5.;
4) Recurrent: Return to Step 2;
5) Stop once all significant inputs have been selected.

**B. ARTIFICIAL NEURAL NETWORK**

Artificial neural network (ANN) are the mathematical tools inspired by the way the human brain processes information. ANN has the high degree of parallel structure and parallel implementation capabilities, and also has the ability to find optimal solutions at high speed. The basic unit of ANN is the artificial neuron, schematically represented in Figure 1. Neurons receive information from multiple input nodes and process it internally to obtain output results. This process typically consists of two phases, first combining the input information linearly and then using the result as an argument to the given activation function [25]. The activation functions represent the nonlinear relationship between the inputs and outputs, which include Sigmoid, Relu functions and so on.

The specific calculation process is as follows:
\[
y_i = f \left( \sum_{i=1}^{N} w_{ij} x_i + b_i \right)
\] (2)
where $y_i$ is output, indicates activation functions, $\omega$ denotes weight, and $x_i$ is input, $b_i$ is bias.

ANN used in various fields is usually composed of many neurons, a typical 3-layer neural network with two hidden layers and one output layer shown in Figure 2. Each layer consists of a set of neurons connected by weight which are randomly initialized and then adjusted by optimization algorithms (e.g. gradient descent and levenberg-marquardt).

The network iteratively adjusts its parameters to reduce the error between the predicted output and the actual output until the error is minimized. The most classic backpropagation neural network (BPNN), that is, signal forward propagation and error backward propagation is applied for forecasting in this paper. And gradient descent approach that the weights are required to be corrected in the direction of the fastest gradient drop as the algorithm of weight update among network. The general calculate steps of weight update are as follows:

**Step 1.** The error between the predicted value $\hat{y}$ and the actual value $y$ is calculated and propagated back.
\[
\delta_k = \frac{1}{2}(y - \hat{y})^2
\]
\[
\delta_k^2 = \omega_k^3 \delta_k
\]
\[
\delta_j^1 = \sum_k \omega_k^2 \delta_k^2
\] (3)

**Step 2.** Adjust the original weight according to the error received in the back propagation process.
\[
\omega_{ij}^1 = \omega_{ij}^1 + \eta \frac{\partial}{\partial \omega_{ij}^1} (e)
\]
\[
\omega_{jk}^2 = \omega_{jk}^2 + \eta \delta_k \frac{\partial (e)}{\partial o_j^2}
\]
\[
\omega_{k}^3 = \omega_{k}^3 + \eta \delta_k \frac{\partial (e)}{\partial o_k^3}
\] (4)

**Step 3.** After the connection weights of each layer of neurons are modified, they enter the next cycle. Then input the next new sample and use the modified weight to forward propagation to get the predicted value. Return to step 1 until
the error value reaches the specified threshold and end the cycle.

There are some problems that cannot be ignored for neural network, such as lowly training speed, easily to fall into local optimum and strong subjectivity. Therefore, on the one hand, genetic algorithm (GA) is used to optimize the initial weights and thresholds of ANN to improve the training speed and performance, as shown in Figure 3. GA is a random search method that draws on the biological evolution law of survival of the fittest. It is also a search heuristic algorithm used to solve optimization in the field of computer science artificial intelligence. On the other hand, the subjective problem of ANN construction is solved by using cross-validation to select the appropriate number of neurons and hidden layers. In the last ten years, ANN has been widely used to predict power load. It is also very well suited for it, for at least two reasons. First, it can approximate numerically any continuous function to the desired accuracy. Second, it is the data-driven approach. That is, the ANN is able to automatically map the relationship between them when given a sample of input and output vectors.

However, the prediction accuracy and training speed of neural networks often depends on whether the appropriate input variables are selected. Electric load demand is affected by many factors, such as weather, economy and special days. Feature selection can reduce the dimension of the input space without sacrificing classification performance. Therefore, a lot of study work begins to focus on features selection before modeling.

III. EXPERIMENT SETTINGS
A. DATA DESCRIPTION
Raw electricity load data used in this study are selected from 0:00:00 on January 1, 2016 to 23:00:00 on December 31, 2018 in Jinan, China, which are collected at hourly time interval. Data from 2016 to 2017 has been used only for forecasting to test the performance of the model on out-of-sample load data. Before modeling the dataset, some preprocessing procedures have been done to enable raw data to become more practical. For example, several missing load values are supplemented by linear interpolation because of that there is only one missing point in every breaking interval. Then we use Pauta standard [34] to identify the abnormal points and treat them as missing values. The pre-processed data is referred to as the original data in what follows, as shown in Figure 4.

B. CANDIDATE FEATURES
Figure 4 illustrates hourly loads from January 1, 2016 to December 31, 2018 in Jinan. The blue curve represents original hourly loads. And daily average load is marked by red curve. It is obvious that load demands have multiple seasonal patterns including the daily and weekly periodicity, especially daily periodicity. The weekly periodicity is only evident in March and April when the power consumption is relatively stable. At the same time, load demand decreased significantly at the weekend. In addition, load levels on national holidays which identified by green curve are lower than on weekdays. This leads us to conclude that load demands are also affected by calendar days. As we all known, holiday load forecasting is a very challenging task because these atypical load conditions are not only rare, but also load variation pattern quite different from normal working days which is caused by the great change in human activities [35]. Therefore, in this study, for the sake of simplicity, we will consider holiday to be similar to weekend, that is weekend and holiday identified by nonworking day and other days by working days. Overall, the power load increases slowly year by year with an average load of 3080.3 MW in 2016, 3096.6 MW in 2017 and 3226.9 MW in 2018, which is consistent with the economic development of Jinan in the past three years. Although the gross domestic product (GDP) of Jinan increases every year, the growth rate is not large at about 7.8%. This also shows the close relationship between the regional GDP level and electricity consumption.

1) SPECTRAL ANALYSIS
According to the research on the internal mechanism of power load data, we know that the load demand is cyclical, the power spectral density of the original load data from 2016 to 2018 is calculated by the Welch’s method for determining the strength of different periodic motion in this time series and shown in Figure 5. The results show that there are three distinct peaks including diurnal, weekly and semidiurnal frequency signal. Among them, the diurnal frequency signal is the dominant component, which is also consistent with the above analysis. Besides that, weekly periodicity also exists in this power load series which is caused by the alternation from working day to nonworking day. It also can be seen from the ordinate of the graph that the intensity of the
2) AVERAGED HOURLY LOAD ANALYSIS OF EACH DAY

Based on the above analysis, we notice that the daily period variation of load demand is the most remarkable. Therefore, data for each day in 2016 and 2017 are averaged and shown in Figure 6. According to the figure, it is obvious that the load varies from hour to hour following the consumers' behavior, and the curves of load data have similar shapes and magnitudes in both years, which indicates that it is necessary to consider daily periodicity in STLF. Moreover, from the change trend of the curve, it can also be inferred that there is a certain relationship between the load of a certain hour and that of the previous several hours.

3) AVERAGED HOURLY LOAD ANALYSIS OF EACH WEEK

Since the weekly periodicity of load demand is also relatively remarkable. Thus, data for each week in 2016 and 2017 are averaged and shown in Figure 7. It is clear that power load on Saturday and Sunday is significantly less than that on weekdays, especially on Sunday. For this problem, a new input feature is added to identify whether the predicted time point belongs to a weekday or weekend, including 0 for non-working days and 1 for working days. What is surprising is that, although the lowest levels of electricity demand in two years are on Sundays, the performance on other days was somewhat different. This is obviously different from average hourly load analysis result of each day in Figure 6. On the
one hand, the weak periodic motion of the cycle obtained by power spectrum analysis is further verified. On the other hand, there may exist a relationship between the load for a given hour on a given day and the load for the same hour in the previous weeks.

4) BOX- PLOT ANALYSIS

Compared with Figure 3, Figure 8 shows the distribution of data in a more abstract way. The blue dotted line represents the average annual load value. The red ‘+’ represents the monthly average load value. It can be seen that compared with summer, electricity consumption level in spring and autumn are more concentrated. And in July and August with the highest temperature throughout the year, the difference between the maximum and minimum load demand is the largest. The existence of this phenomenon will undoubtedly make the prediction of summer become more complicated.

5) CANDIDATE FEATURES

Considering the daily and weekly periodicity characteristics of hourly loads, the hourly load values of the hour of day, the load on the same hour in the previous seven days, previous day’s average load, week of day are selected as the candidate input features of the forecasting model. For the difference between the working day and non-working day load levels, a flag indicating the if it is a weekend/weekday on forecasting time point has had been adopted. On the one hand, there is a common agreement that the temperature is regarded the most important weather influence [25]. On the other hand, due to the complex and diverse topography, which mountainous in the south and Yellow River in the north, and temperate continental monsoon climate, temperature and humidity are two essential factors for load demand in Jinan. As a result, temperature and humidity variables are added for each forecasting time interval, plus the temperature on the same hour in the previous seven days and previous day’s average temperature. By this way, we can consider all the historical data that may have influence on the predicted hour t. Then, the candidate set for model input is summarized as follows:

\[
\text{Candidate-inputs}(t) = \{\text{Hour}(t), \text{Day}(t), W(t), L(t-24), \]
\[L(t-48), \ldots, L(t-144), L(t-168), L-\text{Average}(t), T(t), T(T-24), \]
\[T(t-48), \ldots, T(t-144), T(t-168), T-\text{Average}(t), \text{Humidity}(t)\}. \tag{5}
\]

where t is the time interval index. As hourly load forecasting is studied in this paper, t is on an hourly basis. Where \(L(t-i)\) and \(T(t-i)\) indicates the lagged load and temperature of time interval \(t-i\) respectively, \(\text{Day}(t)\) refers to the day of the week, which is marked by the numbers from 1 to 7. Calendar indicators of hourly are denoted by \(\text{Hour}(t)\), which is marked by the numbers from 1 to 24. \(W(t)\) denotes a flag indicating the if it is a weekend/weekday of \(t\), including 0 for weekends and 1 for weekdays. It is noted that all the public holidays are considered as weekends marked by 0. \(L-\text{Average}(t)\) and \(T-\text{Average}(t)\) indicates the average load of previous day and temperature, respectively. In summary, there are 20 input features in candidate set \(\text{Candidate-input}(t)\).

By removing the candidate features with low relationship between load demand to reduce the size of the input feature set, the prediction engine is able to better learn the input and output mapping function of the process, so as to improve the prediction accuracy and calculate speed. The correlation with candidate inputs above and outputs can be calculated by the filter method based on PMI mentioned in section 2. These candidate features which the PMI value higher than the corresponding 95th percentile value are retained. Feature subset after reduction consist of hour of day, load from the same hour in the previous day, previous day’s average load, day of week, load from the same hour and same day from the previous week, a flag indicating if it is a weekend/weekday, the temperature on the forecasted day and previous day’s temperature, humidity on the forecasted day. Selected inputs are shown as follows:

\[
\text{Selected-inputs}(t) = \{\text{Hour}(t), \text{Day}(t), W(t), L(t-24), L(t-168), \]
\[L-\text{Average}(t), T(t), T(t-24), T-\text{Average}(t), \text{Humidity}(t)\}. \tag{6}
\]
C. PERFORMANCE METRICS

In order to properly evaluate the prediction performance of the proposal, mean absolute percent error (MAPE) accuracy measures method is adopted in this study. The definitions of the method are shown as follows:

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100
\]

where N is the forecasting horizon. This study focuses on the day-ahead short-term load, then the number of forecasting periods N equals 24. And \(y_i\) and \(\hat{y}_i\) represent the actual and predicted loads at period \(i\) respectively. MAPE is a widely used metric that measures the percentage error between actual and predicted values. The smaller the MAPE value, the closer the predicted value is to the actual value, that is, the better the prediction performance of the model is.

D. SELECTED COUNTERPARTS (FOR COMPARISON) AND IMPLEMENTATIONS

To confirm the prediction performance of the proposed feature selection for STLF using ANN, four comparative experiments have been used as counterparts for comparison purposes. These five counterparts are abbreviated as follows:

1. D-ANN: ANN forecasting model considering only daily periodicity.
2. DW-ANN: ANN forecasting model considering daily and weekly periodicity.
3. DWN-ANN: ANN forecasting model considering daily, weekly periodicity and working day/non-working days.
4. DT-ANN: ANN forecasting model considering daily periodicity, temperature and humidity.
5. PMI-ANN: ANN forecasting model with all selected input features.

As a basic experiment, D-ANN model only considers daily periodicity due to daily periodicity is the most remarkable for load demand in Jinan. For all the above-mentioned methods, ANN has been applied as the forecasting model. The detailed input variables and corresponding experiment names for each model are listed in Table 1.

| N0 | Experiment | Model | Input Features |
|----|------------|-------|----------------|
| 1  | Be         | D-ANN | Hour(t), L(t-24), L-Average(t) |
| 2  | Ce-1       | DW-ANN| Hour(t), L(t-24), L-Average(t), Day(t), L(t-168) |
| 3  | Ce-2       | DWN-ANN| Hour(t), L(t-24), L-Average(t), Day(t), L(t-168), W(t) |
| 4  | Ce-3       | DT-ANN| Hour(t), L(t-24), L-Average(t), T(t), T(t-24), T-Average(t) |
| 5  | Ce-4       | PMI-ANN| All features shown in Eq.(6) |

IV. EXPERIMENT RESULTS

Electricity load data measured from 2018 is used for testing the performance of the proposed approach. All simulations are executed in MATLAB environment on the personal computer platform with 2 Intel Core dual-core CPUs (2.4 GHz) and 4 GB memory in Windows 10 environment.

A. COMPARISON FOR EACH HOUR OF THE DAY

The results of five comparative models for forecasting hourly load in 2018 are presented in Figure 9. Which model is represented by the corresponding experiment can be found in Table 1. As can be seen from the figure, the forecasted values of any model are relatively close to the actual ones, which also shows that the basic experiment selected in this paper is reasonable. However, it is undeniable that the predicted value between 3 pm and 6 pm is very different.

B. COMPARISON FOR EACH DAY OF THE WEEK

The comparison between actual and forecast average load using different models for each day of the week is presented in Figure 10. Among them, the black curve represents the actual load, and the meaning of the other curves is the same as that shown in Figure 9. For sake of simplicity, it will not be repeated. From the results, the following conclusions can also be drawn:

(a) Compared with D-ANN and DT-ANN models without considering week periodicity, the simulation results of other models are relatively close to the actual values.

(b) On Saturday and Sunday, DWN-ANN performed best among five models. In addition, it can be seen from the figure that there is a large gap between the simulation results on Saturday, which is caused by the shift from working days to non-working days.

(c) It is worth noting that PMI-ANN performs better on weekdays and Sundays, and the error on Saturdays is not too large. Unlike DT-ANN, although the accuracy rate is high on non-working days, it is too low on working days to meet the daily forecast requirements of the power system.

Through comprehensive consideration of the above results, PMI-ANN is much better than any other model in the actual forecasted week of 2018 in Jinan.
than other counterparts. This is because other models that consider single or a few features have larger errors in more months, while PMI-ANN only has higher error values in February. For instance, the simulation errors of D-ANN and DW-ANN are relatively high in most months of 2018. Because they consider few factors, they cannot fully capture the law of load changes. But it must be admitted that the DT-ANN model has a very low and stable simulation error value every month, second only to the PMI-ANN model. This further illustrates the important influence of temperature on electricity consumption in Jinan.

V. DISCUSSION

According to these simulation results presented in section 4, we can conclude that temperature are the most important factors for power load variation in Jinan which is mainly due to topographic and climatic factors. Jinan is surrounded by mountains on three sides, north by Yellow River, among them, the influence of mount Taishan mountain forms the foehn effect. The south wind is easy to form the sinking foehn wind, and the cold air from the north is also easy to enter. In addition to the narrow pipe effect of the terrain, the cold and hot air are not easy to dispersed. Consequently, Jinan city is a typical southerly heat island and a cold wave island. On southerly nights, Jinan’s high night temperature is the most obvious in Shandong province, but when the cold air comes from the north, it will not only have no resistance but also form a accumulation, leading to colder. Furthermore, the amplitude of winter and spring air temperature in Jinan is rare in the whole country. In view of the above specific regional characteristics and experimental results, our future research work will further concentrate on the study of the huge influence of temperature to improve the prediction accuracy of a given month.

The second major factor affecting the load is the period of days, which is verified by the above experimental results. In summary, the ANN forecasting model with selected features by filter method based on PMI significantly receive the relatively high forecasting accuracy in terms of MAPE. At the same time, the forecasting performance of these model with single factors are also not as good as PMI-ANN model. This phenomenon is mainly due to few input features make model unable to accurately simulate the complex relationship between inputs and output.

Although the proposed method has only been trained and tested on the load data measured from 2016 to 2018 in Jinan, it is also applicable to the future load demand of Jinan and other areas. This is because the features selected through the feature selection are the most suitable for Jinan, and climate, topography, and economic conditions in Jinan will not change dramatically in a short time. Therefore, the proposed approach in this paper can be used for future load demand forecasting in Jinan. For other areas, reselection of candidate features is necessary when the simulation results obtained by using the candidates extracted in this paper are not ideal. The characteristics of load changes in each region are different, but excellent results should be obtained based on the ideas in

C. COMPARISON FOR EVERY MONTH

Figure 11 shows the MAPE error of all different models for every month in 2018, it is interesting to note that DT-ANN and PMI-ANN considering temperature as a input present significantly lower MAPE, especially during hot summer and cold winter, which also shows that the temperature is more important in forecasting the load than other lagged load. This is mainly due to the climate and topography of Jinan, which will be described in detail in the discussion section.

In addition, we can conclude that PMI-ANN has a lower MAPE value in all comparison models from the observation and analysis of the annual data. And, there is the lowest error value in the eight months of the year. It is undeniable that the error value in February is higher than other models, which is caused by national holidays.

We can also observe that the MAPE error of any model in February and summer is significantly larger than that in other months. However, the errors of all models are relatively low and similar throughout the spring. As is known to all, Chinese lunar spring festival, which is stipulated as a national legal holiday and whose holiday is set at the end of January and the beginning of February, is the most important traditional holiday of the year for Chinese people. And summer is the time when most of students have their holidays. Besides, the temperature rises rapidly from spring to summer and high temperatures in summer are a unique characteristic of Jinan. Therefore, load trend of these month are slightly different to normal days and bring some difficulties for STLF.

In a word, from the MAPE value of each month in Jinan throughout the year, the PMI-ANN model performs better than other counterparts. This is because other models that
this paper. After work, we applied the proposed approach to other cities of Shandong province to verify the assumption.

VI. CONCLUSION

To simplify the learning process of forecasting models to reduce running time and better simulate the nonlinear relationship between load and relevant factors to improve prediction accuracy, feature selection is an important stage in STLF. In this study, we newly proposed a multifactorial framework which composed of data analysis, PMI-based filter method and ANN to address the problem. At the same time, we implement a graphical tool for easy and accurate computation of day-ahead system power load forecast with MATLAB App Designer. The performance of the proposed approach is tested on data in Jinan, and the following main conclusions are drawn from the simulation results:

(1) Through detailed data analysis such as power spectrum analysis, the main periodic movement of the load is found, and candidate features with good performance are extracted.

(2) The PMI-based filter which can easy to implement and fast in calculation is used for feature selection, and the most classic BP neural network is adopted to enable the simulation process to achieve faster calculation speed while maintaining accuracy.

(3) Five comparative experiments with different features are designed and implemented, and the results show that the selected features using the proposal are better than the results using single or a few of features.

(4) As analyzed by the above experimental data, it can be seen from Figure 9 that the periodic characteristics of the day have a better effect on STLF. It can also be clearly seen from Figure 11 that temperature has a greater impact on changes in load demand, so for Jinan, temperature is a factor that must be considered in load forecasting. At the same time, it further illustrates that the proposed feature selection method can accurately extract influencing factors.

The MF proposed in this paper can be used not only for electricity load demand forecasting, but also for electricity price forecasting, image recognition and so on.

REFERENCES

[1] E. Ceperic, V. Ceperic, and A. Baric, “A strategy for short-term load forecasting by support vector regression machines,” IEEE Trans. Power Syst., vol. 28, no. 4, pp. 4356–4364, Nov. 2013.

[2] H. Hahn, S. Meyer-Nieberg, and S. Pickl, “Electric load forecasting methods: Tools for decision making,” Eur. J. Oper. Res., vol. 199, no. 3, pp. 902–907, Dec. 2009.

[3] G. Gross and F. D. Galiana, “Short-term load forecasting,” Proc. IEEE, vol. 75, no. 12, pp. 1558–1573, Dec. 1987.

[4] E. Gonzalez, M. A. Jaramillo and D. Carmona, “Monthly electric energy demand forecasting based on trend extraction,” IEEE Trans. Power Syst., vol. 21, no. 4, pp. 1946–1953, Dec. 2006.

[5] W. Charytoniuk, M. S. Chen, and P. Van Olinda, “Nonparametric regression based short-term load forecasting,” IEEE Trans. Power Syst., vol. 13, no. 3, pp. 725–730, Aug. 1998.

[6] K. B. Song, Y. S. Baek, D. H. Hong, and G. Jang, “Short-term load forecasting for the holidays using fuzzy linear regression method,” IEEE Trans. Power Syst., vol. 20, no. 1, pp. 96–101, Feb. 2005.

[7] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis Forecasting and Control, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.

[8] S. J. Huang and K. R. Shih, “Short-term load forecasting via ARMA model identification including non-Gaussian process considerations,” IEEE Trans. Power Syst., vol. 18, no. 5, pp. 673–679, Jul. 2003.

[9] X. Y. Xu, Y. Q. Qi and Z. S. Hua, “Forecasting demand of commodities after natural disasters,” Expert Syst. Appl., vol. 37, pp. 4313–4317, Dec. 2012.

[10] G. Cecati, J. Kolbusz, P. Siano, and B.M. Wilamowski, “A novel RBF training algorithm for short-term electric load forecasting and comparative studies,” IEEE Trans. Ind. Electron., vol. 62, no. 10, pp. 6519–6529, Dec. 2015.

[11] H. Yoo and R. L. Pimmel, “Short term load forecasting using a self-supervised adaptive neural network,” IEEE Trans. Power Syst., vol. 14, no. 2, pp. 779–784, May 1999.

[12] D. Park, M. Shkarawki, I. R. J. Marks, L. Atlas, and M. Damborg, “Electric load forecasting using an artificial neural network,” IEEE Trans. Power Syst., vol. 6, no. 2, pp. 442–449, May 1991.

[13] H. S. Hippet, D. W. Bunn, and R. C. Souza, “Large neural networks for electricity load forecasting: Are they overfitted?” Int. J. Forecasting, vol. 21, no. 3, pp. 425–434, Jul. 2005.

[14] H. J. Sadaei, F. G. Guimares, C. J. O. Silva, M. H. Lee, and T. Eslami, “Short-term load forecasting method based on fuzzy time series, seasonality and long memory process,” Int. J. Approx. Reasoning, vol. 83, pp. 196–217, Apr. 2017.

[15] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, “Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings,” Appl. Energy, vol. 195, pp. 659–670, Jun. 2017.

[16] G.-F. Fan, L.-L. Peng, and W.-C. Hong, “Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model,” Appl. Energy, vol. 224, pp. 13–33, Aug. 2018.

[17] R.-A. Hooshmand, H. Amooshahi, and M. Parastegari, “A hybrid intelligent algorithm based short-term load forecasting approach,” Int. J. Electr. Power Energy Syst., vol. 45, no. 1, pp. 313–324, Feb. 2013.

[18] M. Niu, S. Sun, J. Wu, L. Yu, and J. Wang, “An innovative integrated model using the singular spectrum analysis and nonlinear multi-layer perceptron network optimized by hybrid intelligent algorithm for short-term load forecasting,” Appl. Math. Model., vol. 40, nos. 5–6, pp. 4079–4093, Mar. 2016.

[19] Z.-B. Shi, Y. Li, and T. Yu, “Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm,” in Proc. Int. Conf. Inf. Multimedia Technol., 2009, pp. 306–309.

[20] Z. Zhang, W.-C. Hong, and J. Li, “Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm,” IEEE Access, vol. 8, pp. 14642–14658, 2020.

[21] O. Abedinia, N. Amjady, and H. Zareipour, “A new feature selection technique for load and price forecast of electrical power systems,” IEEE Trans. Power Syst., vol. 32, no. 1, pp. 62–74, Jan. 2017.

[22] M. Rana, I. Koprinska, and A. Khotavsi, “Feature selection for interval forecasting of electricity demand time series data,” in Artificial Neural Network, vol. 4. New York, NY, USA: Springer, 2015, pp. 445–462.

[23] S. Kouhi, F. Keynia, and S. Najafi Ravandanegh, “A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection,” Int. J. Electr. Power Energy Syst., vol. 62, pp. 862–867, Nov. 2014.

[24] S. Li, P. Wang, and L. Goel, “A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection,” IEEE Trans. Power Syst., vol. 31, no. 3, pp. 1788–1798, May 2016.

[25] H. S. Hippet, C. E. Pedreira, and R. C. Souza, “Neural networks for short-term load forecasting: A review and evaluation,” IEEE Trans. Power Syst., vol. 16, no. 1, pp. 44–55, Dec. 2001.

[26] Y. Li and Z. Yang, “Application of ESO-ELM with binary jaya-based feature selection to real-time transient stability assessment using PMU data,” IEEE Access, vol. 5, pp. 23002–23011, 2017.

[27] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, “Hybrid filter–wrapper feature selection for short-term load forecasting,” Eng. Appl. Artif. Intell., vol. 40, pp. 17–27, Apr. 2015.

[28] N. Amjady and F. Keynia, “A new neural network approach to short term load forecasting of electrical power systems,” Energies, vol. 4, no. 3, pp. 488–503, Mar. 2011.

[29] C. L. Wu and K. W. Chau, “Data-driven models for monthly streamflow time series prediction,” Eng. Appl. Artif. Intell., vol. 23, no. 8, pp. 1350–1367, Dec. 2010.
[30] F. Keynia, “A new feature selection algorithm and composite neural network for electricity price forecasting,” Eng. Appl. Artif. Intell., vol. 25, no. 8, pp. 1687–1697, Dec. 2012.

[31] A. Sharma, “Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part I—A strategy for system predictor identification,” J. Hydrol., vol. 239, nos. 1–4, pp. 232–239, Dec. 2000.

[32] I. Luna, S. Soares and R. Ballini, “Partial mutual information criterion for modelling time series via neural networks,” in Proc. 11th Inf. Process. Manage. Uncertainty Int. Conf., 2006, pp. 1–5.

[33] G. J. Bowden, G. C. Dandy and H. R. Maier, “Input determination for neural network models in water resources applications. Part 1—Background and methodology,” J. Hydrol., vol. 301, pp. 75–92, Dec. 2005.

[34] H. Zhao, F. Min, and W. Zhu, “Test-cost-sensitive attribute reduction of data with normal distribution measurement errors,” Math. Probl. Eng., vol. 3, pp. 4928–4942, Dec. 2013.

[35] P. Zeng, C. Sheng and M. Jin, “A learning framework based on weighted knowledge transfer for holiday load forecasting,” J. Mod. Power Syst. Clean Energy, vol. 7, pp. 329–339, Dec. 2019.

[36] H. Dong, Y. Gao, X. Meng, and Y. Fang, “A multifactorial short-term load forecasting model combined with periodic and non-periodic features—A case study of Qingdao, China,” IEEE Access, vol. 8, pp. 67416–67425, 2020.

**YAN GAO** is currently pursuing the degree with the College of Mathematics and Systems Science, Shandong University of Science and Technology. Her research interests include electricity load forecasting and artificial intelligence.

**YONG FANG** received the B.S. degree in mathematics and applied mathematics and the M.S. degree in basic mathematics from the Shandong University of Science and Technology, in 2008 and in 2011, respectively, and the Ph.D. degree in physical oceanography from the Institute of Oceanology, Chinese Academy of Sciences, in 2014. From 2014 to 2018, he worked with the State Key Laboratory of Tropical Marine Environment, South China Sea Institute of Oceanology, Chinese Academy of Sciences. Since June 2018, he has been a Lecturer with the College of Mathematics and Systems Science, Shandong University of Science and Technology. He has been engaged in applied oceanography, power big data and other related research work, presided over or participated in more than 20 scientific research projects, and published more than ten high-level academic articles.

**HUANHE DONG** received the B.S. degree in mathematics education from Qufu Normal University, in 1994, and the M.S. degree in applied mathematics and the Ph.D. degree in computer software and theory from the Shandong University of Science and Technology, in 2000 and in 2008, respectively. From 1994 to 1998, he was a Teaching Assistant with the College of Science, Shandong University of Science and Technology. From 1999 to 2003, he was a Lecturer with the College of Information Science and Engineering, Shandong University of Science and Technology, where he was an Associate Professor, from 2004 to 2008, and a Professor, from 2008 to 2014. Since 2014, he has been a Professor with the College of Mathematics and Systems Science, Shandong University of Science and Technology. He is currently a Professor, a Ph.D. Tutor, and the Dean of the College of Mathematics and System Science, Shandong University of Science and Technology. He is the author of three books and more than 40 articles. His research interests include data analysis, stability of differential equations, and solitary theory and its application. He received several awards and honors, including the third prize of the Natural Science Award of Shandong Province, the second prize of the Qingdao Natural Science Award, and so on.

**YUAN KONG** received the M.S. degree in computational mathematics from the Shandong University of Science and Technology, in 2011, and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, in 2015. From 2015 to 2017, she worked with South China Agricultural University. Since January 2018, she has been a Lecturer with the College of Mathematics and Systems Science, Shandong University of Science and Technology. She has been engaged in biological computing, power big data, and other related research work, presided over or participated in more than ten scientific research projects, and published more than ten high-level academic articles.

***