

Received October 17, 2020, accepted November 1, 2020, date of publication November 5, 2020, date of current version November 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036096

An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend

MD FAZLA ELAHE^{ID}, MIN JIN^{ID}, (Member, IEEE), AND PAN ZENG^{ID}, (Associate Member, IEEE)

College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

Corresponding author: Min Jin (jinmin@hnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773157, in part by the Hunan Provincial Innovation Foundation for Postgraduate under Grant CX2018B231, and in part by the Changsha Key Research and Development Program under Grant KQ2004011.

ABSTRACT Modeling an accurate forecasting model for short-term load is still challenging due to the diverse causes of load changing and lack of information on many of these causes. In this paper, error trend is used to reveal the trend effect caused by unknown load affecting factors and proposed adaptive second learning of error trend (A-SLET) to self-adapt the trend effect. Furthermore, the training set is classified based on balance point temperature and then parallelly trained and tested adaptive forecaster for hot days and adaptive forecaster for cold days with proper data. Combining A-SLET with parallel forecasting and training set classification, Adaptive and Parallel forecasting strategy based on Second Learning of Error Trend (AP-SLET) is proposed. The work studied two distinct load patterns, one in the USA and the other in Australia. Considering the yearly forecasting horizon, MAPE of the adaptive and parallel forecasting strategy is 1.87%-4.04% for ME-Maine of New England and 2.81%-4.41% for New South Wales. Compared to the state-of-art forecasting methods, MAPE of the adaptive and parallel forecasting strategy is reduced by 17.03%-33.33%, RMSE and MAE are reduced by 34.05% and 35.38% respectively. The experimental results demonstrate the proposed strategy can transform unknown and unavailable load affecting factors into known forecasting features and then adapt it to improve forecasting performance. The proposed strategy is also forecaster independent and equally applicable to almost all load scenarios regardless of geographical and seasonal differences.

INDEX TERMS Adaptive and parallel forecasting, short-term load forecasting, smart grid, second learning of error trend, training set classification.

I. INTRODUCTION

Building an error-free load forecasting model is challenging due to the diverse use of electricity and various random and non-random factors. Among energy conscious users, the widespread use of energy-efficient appliances, and the use of different renewable energy sources are making load forecasting more complex.

Up to now, numerous studies have been conducted to tackle the problem of short-term load forecasting and smart grid management. The most frequently used models in the literature are the statistical models that try to find the qualitative relation between the historical load data and future load in a time series. The auto-regressive model and its

variations [1] are widely employed statistical models. Some other statistical models are the gray model [2], multiple linear regression [3], etc. Although most of the statistical models require less computational resources, they consider only linear relationships and can only be applied when dealing with relationships that graphically look like a straight line. Researchers also introduced several soft computing techniques such as the improved fuzzy model [4], expert system [5], etc. Knowledge-based methods [6], [7] are also studied for load forecasting. Despite of the fact that these soft computing models handle non-linear behavior and do not require many computational resources, the accuracy is not up to the expectation.

To adapt the non-linear behavior of various load changing factors and to improve the accuracy, researchers showed their interest in artificial intelligence [8]. The most

The associate editor coordinating the review of this manuscript and approving it for publication was Dongxiao Yu^{ID}.

popular such models are artificial neural network [9], support vector machine [10], decision tree [11], etc. In recent years deep learning based techniques such as Long Short-Term Memory (LSTM) [12], Gated Recurrent Units (GRU) [13] gain popularity. Several algorithms such as particle swarm optimization [14], multi-objective dragonfly algorithm [15], follow the leader [9], ant lion optimizer [16], firefly algorithm [17], artificial bee colony [17], and grasshopper optimization algorithm [10] proposed for solving the parameters and hyperparameters of these models. Later on, researchers found the individual forecasters may suffer from uncertainty, high sensitivity to model parameters, and over-fitting problems.

To mitigate the above-mentioned problems caused by the single forecaster, researchers proposed ensemble technique [18]–[23]. In the ensemble model, several independent forecasters are combined to improve the robustness over single forecaster. Another crucial benefit of using ensemble techniques is that it improves accuracy and can be used as a sophisticated tool for forecasting. Although the ensemble models ensure high accuracy, the accuracy still depends on the inclusion of load affecting factors [22]. The more the load affecting factors included in the forecasting model, the higher the accuracy will be. But in reality, it is quite difficult to know all the load affecting factors.

Feature selection is the process of selecting the most contributing features for the output. Whereas, training set construction is the technique to select particular data that reflects the pattern of the load variation. To increase the accuracy of load forecasting, training set construction plays an important role. Training set construction rarely explored, compared to feature selection and model development. Researchers found that the training set containing similar days as the forecasted days trains the model better. There are several ways [24]–[29] to find similar days from the historical load data. Although the same method is used for finding similar days, the number, variable type, and weight factors are different from each other. Reference [27] introduced the index-mapping database and proposed an improved similar days method. However, the source of mapping values was not described. A day-to-day topological network considering feature similarity of historical day and forecasted day proposed in [30], where all the features are considered with the same importance to generate the day-to-day network. A recent research article [10] also used the traditional similar day approach to construct training set, considering temperature and humidity. Since the load affecting factors change over time, it is not wise to select a training set which has a long gap from the forecasted day. The impact of calendar effects and forecast granularity for short-term residential load forecasting examined in [31]. Even though the use of similar day approach increases accuracy, proper training of the model using this approach requires a large amount of historical data. It is also computationally expensive if more than one variable is considered for finding similar days.

Aiming to solve the above-mentioned challenges, A-SLET is newly proposed. First, important load changing factors are

combined to construct a workable dataset for our load forecasting model. Next, the test set is excluded from the learning set. The learning set is further divided into three sets and eventually, the continuous validation set is obtained. After that, the error of the validation set is measured. To get the error trend, curve fitting is applied to the smooth error signal of the continuous validation set. The purpose of using the error trend is to overcome the lack of information and reveal the trend effect caused by unknown load affecting factors. After that, the error trend along with the original features is used for second learning. The purpose is to self-adapt the trend effect caused by unknown load affecting factors. To avoid the problem of training set construction discussed earlier, the training set is divided into two different sets: cold days and hot days based on balance point temperature [3]. A parallel training and forecasting are applied to get the forecasting result. Finally, AP-SLET is proposed combining training set classification, parallel forecasting, and A-SLET. In this paper, A-SLET and AP-SLET are applied to forecast the daily peak load of two public datasets with vast geographical differences. The effectiveness of A-SLET and AP-SLET is explored on distinct patterns of load by performing experiments and comparative studies.

The rest of this paper is structured as follows. Section II and section III introduce A-SLET and AP-SLET respectively. Section IV provides experimental results and associated discussions. The concluding remarks are given in Section V.

II. ADAPTIVE FORECASTING STRATEGY BASED ON SLET

Second learning [22] sometimes refers to stacking approach [32], [33], is a common technique for merging the results of more than one machine learning model. Suppose (s_1, s_2, \dots, s_n) and (g_1, g_2, \dots, g_n) are the generated results of models S and G respectively. To merge the results of S and G, another model M is trained and forecasted using the generated results of S and G. However, our proposed SLET is quite different from the conventional approach. In the conventional approach, the generated result of same time horizon is used as input to another model but in our model, it is not the same time horizon. Moreover, error trend is estimated from the generated result, and later on another model is trained and tested using the original data along with the error trend. Error trend presents the estimated random variation free error of the specific time horizon. By incorporating the error trend, the accuracy of the forecaster can be improved. The error of forecaster is mainly caused by some unconsidered load affecting factors. So, the error trend analysis allows us to incorporate the trend of unknown load affecting factors in our forecasting strategy. It also enables us to consider the unavailable affecting factors.

A. DATASET CONSTRUCTION AND DIVISION

The accuracy of electricity load forecasting depends on the inclusion of important load changing factors. Generally, the electricity load dataset does not contain information on different load changing factors. In this paper, weather,

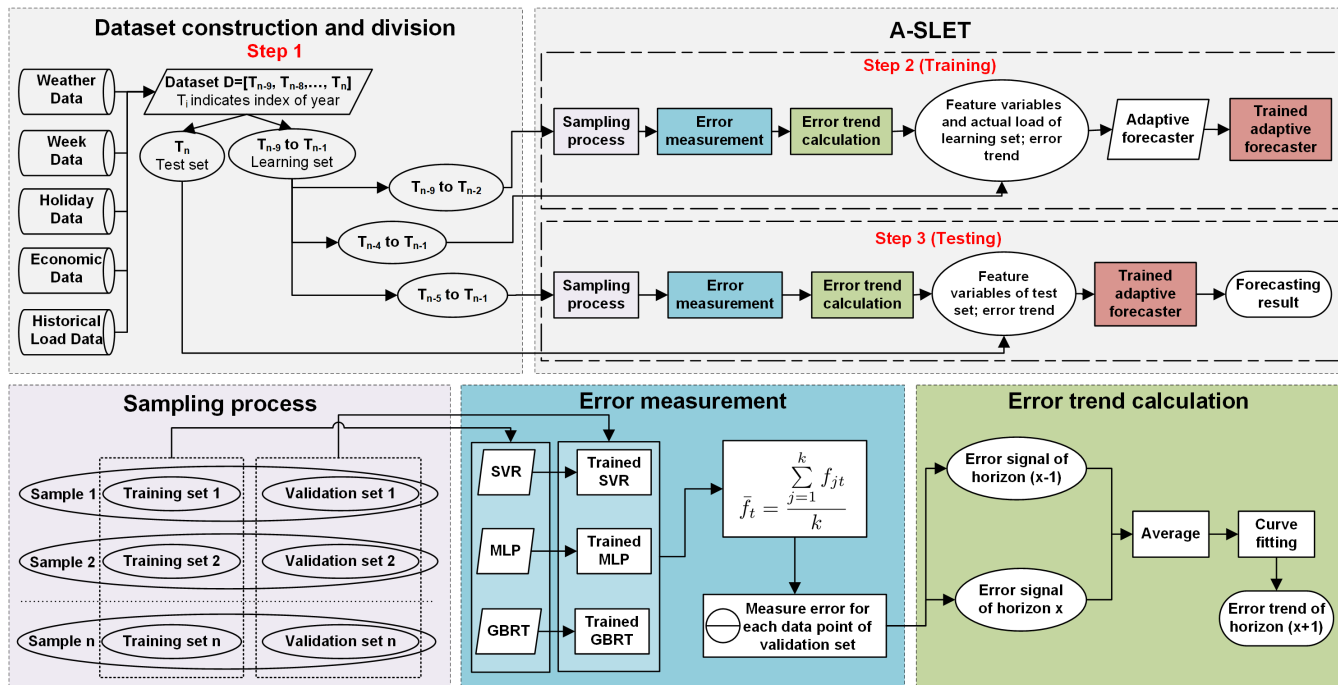


FIGURE 1. Adaptive forecasting strategy based on second learning of error trend.

calendar, economic are considered as important load changing factors and combined with load data to maximize accuracy. Suppose our dataset $D = [T_{n-9}, T_{n-8}, \dots, T_n]$, where T_i indicates index of year. Each year has 365 days so each index of year T_i has 365 time-lag. Each time lag has features of important load changing factors and load for the specific time lag. The first step of getting error trend is the sampling process and the procedure is demonstrated in Fig. 1. To do so, the whole dataset is divided into a learning set (T_{n-9} to T_{n-2}) and a test set (T_n). Later on, the learning set is divided into following three parts if three years of data as the training set for error measurement and four years data as second learning training set is considered: T_{n-9} to T_{n-2} , T_{n-5} to T_{n-1} and T_{n-4} to T_{n-1} . The need for such division and flow of data are demonstrated in Fig. 2. The three data parts are for getting the error trend for training set, getting error trend for testing set and second learning training set itself respectively.

B. SAMPLING PROCESS

The first two parts of the learning set go through the sampling process and several samples each of which consisting a training set and a validation set are generated. The samples are constructed in such a way that the validation sets become continuous. For example, if the first sample has the training set of $T_{n-4}, T_{n-3}, T_{n-2}$, and validation set of T_{n-1} then the second sample should have training set of $T_{n-3}, T_{n-2}, T_{n-1}$ and validation set of T_n . The reason for the continuous validation set is to get the error trend for the training set and test set.

C. ERROR MEASUREMENT

The second step is the error measurement of continuous validation set. As discussed earlier, a single forecaster suffers

from limited generalization ability, high sensitivity to model parameters, and overfitting problems. The accuracy of a single forecaster fluctuates excessively if applied to different forecasting instances. On the other hand, ensemble models are relatively stable with better accuracy. Considering the above benefits, studies suggested using the ensemble model. Although the ensemble model is proposed to use, our proposed techniques are model-independent, which means the proposed techniques can be used with the model to improve its accuracy. The three forecasters that are used in our proposed framework have their advantage and disadvantage. But together they can build a robust model [22]. For our proposed techniques, the forecasters used to construct ensemble mode are support vector regressor (SVR) [34], multilayer perceptron (MLP) [35], and gradient boosting regression tree (GBRT) [22]. The choice of the forecaster is inspired by the suitability of the forecaster for time series analysis, parameter optimization, and running time.

There are several techniques available for merging the results of individual forecaster such as arithmetic averaging, regression-based merging, performance-based merging, etc. In this research paper, the simple average technique used for merging the results forecasted by the individual forecaster. For K forecasters, forecasting results $f_{1t}, f_{2t}, \dots, f_{kt}$ at the time t , the merged load is given by,

$$\bar{f}_t = \frac{\sum_{i=1}^k f_{it}}{k}$$

where \bar{f}_t denotes the merged load at the time t . The error of the validation set is calculated by the difference between the actual load in the validation set and the forecasted load.

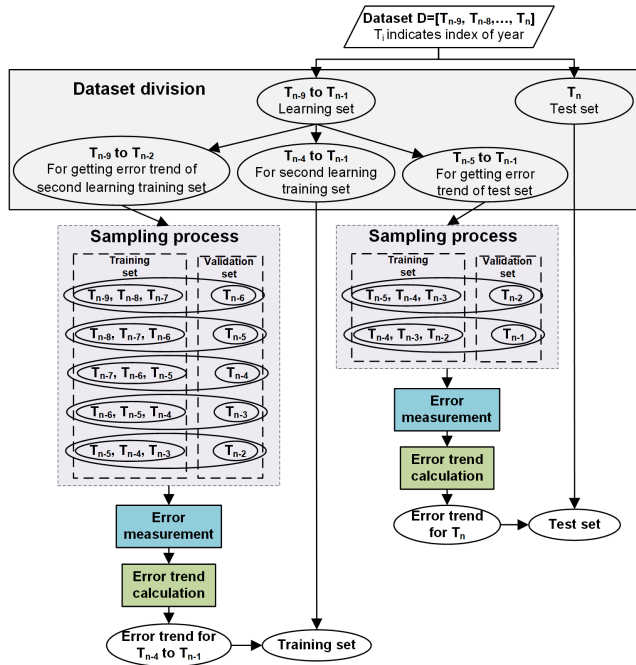


FIGURE 2. Dataset division and flow of data.

D. ERROR TREND

As the error signal of the continuous validation set is obtained, the third step is to calculate the error trend. The process for obtaining the error trend is also given in Fig. 1. The average error signal of two consecutive years is considered for curve fitting and the reason behind it is to use a relatively stable error signal for curve fitting. Hence, the error trend produced by the process is considered as the error trend of next horizon. For example, the error signal of the year T_{n-1} and T_n is used to produce the error trend of T_{n+1} . Curve fitting is proposed to produce error trend from two error signals. It is the process of constructing a mathematical function that best fits the series of given data points. The constructed mathematical function allows us to create new data points that can represent the original curve. Hence, polynomial interpolation refers to the interpolation by the polynomial of the lowest possible degree. Suppose that the polynomial interpolation is in the following form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 \quad (1)$$

The statement that p interpolates the data points means that

$$p(x_i) = y_i \quad \text{for all } i \in \{0, 1, 2, \dots, n\}$$

Equation (1) can be represented as follows,

$$\begin{bmatrix} x_0^n & x_0^{n-1} & x_0^{n-2} & \dots & x_0 & 1 \\ x_1^n & x_1^{n-1} & x_1^{n-2} & \dots & x_1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_n^n & x_n^{n-1} & x_n^{n-2} & \dots & x_n & 1 \end{bmatrix} \begin{bmatrix} a_n \\ a_{n-1} \\ \dots \\ a_0 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{bmatrix}$$

The solution minimizes the following squared error so that the newly constructed curve becomes the best-fitted curve in

the lowest possible degree.

$$E = \sum_{i=0}^n |p(x_i) - y_i|^2$$

Curve fitting is required when a curve fluctuates in an excessive manner, and a smooth curve is needed that can represent the original curve. It removes the random variation and shows the trend and cyclic components. In this paper, curve fitting is used to get the smooth and random variation free error trend.

E. SECOND LEARNING AND ADAPTIVE STRATEGY

The error trend along with the existing features are then used to train the adaptive forecaster. Hence, the error adaption is done by second learning because first learning is used to generate the error trend. Equation (2) and (3) show the way to include the error trend in the training set and test set for the adaptive forecaster. The adaptive forecaster is basically a forecasting model whose function is to adapt the error trend in second learning. The inclusion of the error trend in training set and test set makes the strategy adaptive enabling adaptive forecaster to adapt the effect of unknown load changing factors. There are n numbers of features and m instances present in the training set. f_i^j , E_i and L_i present feature value of j, error trend value, and actual load value at the time i respectively. Among them L_i is the target variable and only used in the training set. Each instance in the training set can be any forecasting granularity. For our this experiment, we used the daily forecasting granularity is used. The full procedure of A-SLET is presented in Fig. 1. We propose to use multi-source data for improving forecasting accuracy. As the name suggested, information fusion of data from multiple sources constructs a multi-source dataset. Generally, the electricity load dataset does not contain information on different load changing factors. So, to include the important load changing factors, it is necessary to consider multi-source data for improving the accuracy of electricity load forecasting. In this paper, weather, calendar, economic and historical load data are considered as multi-source data.

$$Train = \begin{bmatrix} f_1^1 & f_1^2 & f_1^3 & \dots & f_1^n & E_1 & L_1 \\ f_2^1 & f_2^2 & f_2^3 & \dots & f_2^n & E_2 & L_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ f_m^1 & f_m^2 & f_m^3 & \dots & f_m^n & E_m & L_m \end{bmatrix} \quad (2)$$

$$Test = \begin{bmatrix} f_{m+1}^1 & f_{m+1}^2 & \dots & f_{m+1}^n & E_{m+1} \\ f_{m+2}^1 & f_{m+2}^2 & \dots & f_{m+2}^n & E_{m+2} \\ \dots & \dots & \dots & \dots & \dots \\ f_{m+k}^1 & f_{m+k}^2 & \dots & f_{m+k}^n & E_{m+k} \end{bmatrix} \quad (3)$$

III. ADAPTIVE AND PARALLEL FORECASTING STRATEGY BASED ON SLET

The necessity of constructing a training set is already discussed in section I. In this paper, a new method of constructing the training set is proposed, that uses the concept of training set classification. The proposed training set classification method relies only on a single variable that is temperature. As it relies only on a single variable, it does not need any

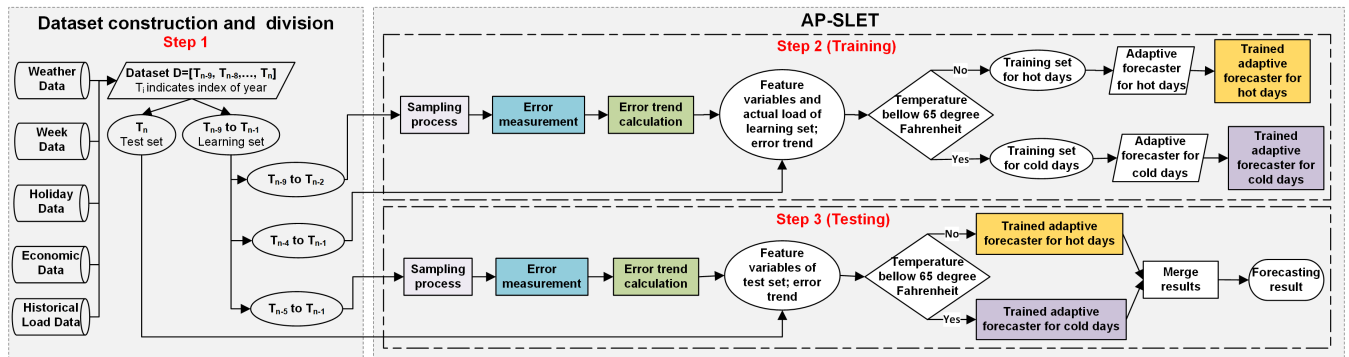


FIGURE 3. Adaptive and parallel forecasting strategy based on second learning of error trend.

weight factors. Another advantage of the proposed method is that it can turn some features to linear, which means the demand of electricity follows a straight line with the increase or decrease of the linear feature. This may help the linear forecasting model to be more accurate.

A. BALANCE POINT TEMPERATURE

The point of temperature, which is neither too hot nor too cold, is called the balance point temperature. Both the training set and test set can be divided based on the balance point temperature. The need for electricity in the balance point temperature is minimum because of the minimum electricity requirement for heating and cooling purpose. Most of the literature considers 65°F as a balance point temperature (e.g. [3]). In this experiment, days with the temperature below to the balance point temperature are called cold days, and days with the temperature above or equal to the balance point temperature are called hot days. Based on the balance point temperature, a new training set classification method is proposed. In the proposed method, the training set is divided into cold days and hot days for the proper training of the adaptive forecaster.

B. PARALLEL FORECASTING

The parallel forecasting is the type of forecasting where multiple forecasters are trained and forecasted simultaneously [19]. The long-time horizon is divided into several horizons, that can be forecasted parallelly. The final forecasting result is produced by merging the results of multiple forecasters. Suppose year index T_n to forecast, which may be considered as long-time horizon. As each year has 365 days so, there are 365 time-lags to forecast if daily peak load is considered. The 365 time-lags can be divided into $((t_1, t_{60}), (t_{61}, t_{100}), \dots, (t_{320}, t_{365}))$, which may be considered as short-time horizon. The division of the time-lag is user-dependent and requires domain knowledge to maximize accuracy of the forecaster. The forecaster $f_1, f_2, f_3, \dots, f_n$ can be trained to forecast the divided small horizon parallelly. The parallel forecasting allows us to use different training sets to forecast specific test sets. For a distinct type of test set,

parallel forecasting is highly effective. In our experiment case, the training set is divided into hot days and cold days based on the balance point temperature. Similarly, the forecasted horizon is divided into hot days and cold days as well. Two adaptive forecasters are then trained using cold days and hot days to forecast the cold days and hot days in the forecasted horizon.

C. PROPOSED FORECASTING STRATEGY

Combining A-SLET, the training set classification, and parallel forecasting, a new strategy for load forecasting called AP-SLET is proposed. The flow chart of AP-SLET is shown in Fig. 3 and the process of sampling, error measurement, and error trend calculation are shown in Fig. 1. The whole process starts with dividing the dataset into a learning set and a test set. Sampling process, error measurement, and error trend calculation procedures are then applied to the corresponding part of the learning set. The training set for second learning is divided into cold days and hot days based on balance point temperature. Two adaptive forecasters are then parallelly trained using two proper training sets. Cold days adaptive forecaster and hot days adaptive forecaster are trained with the cold days training set and hot days training set. The trained adaptive forecasters are used to forecast cold days and hot days of the test set. The final forecast is made by merging the results of two adaptive forecasters.

IV. EXPERIMENTAL STUDY

A. DATASET PREPARATION

The proposed forecasting strategies are applied to two different datasets with a large geographical area and distinct load patterns to show effectiveness. One of the datasets is from the New England region of the USA [36], and another is from New South Wales (NSW) of Australia [37]. ISO New England Inc. is responsible for reliable operation of New England's electric power generation and transmission system. In this experiment, the daily peak load of Maine of New England is used. Load data of New South Wales are collected from the Australian Energy Market Operator. Meteorological data are obtained from [38]. To experiment with the ME peak

load, 23 meteorological features, 9 calendar features, and one economic feature are collected from multiple sources. NSW data include 12 weather features, 7 calendar features, and the price of the electricity. The considered features are collected from multiple sources and merged to make a workable dataset. However, the merging depends on the availability of these features for the specific load zone in the required frequency. The experiment datasets are noise-free and there are no empty values. This allows us to skip the data preprocessing part.

B. EXPERIMENTAL ENVIRONMENT SETUP

The experimental environment includes the Intel (R) Core (TM) i5-8250U CPU (1.80GHz, 8GB memory), and the operating system is Windows 10 (64-bit). Five widely used forecasters: SVR, GBRT, MLP, LSTM and GRU, which are capable of coping with the nonlinear relationship of features, are used to show the effectiveness of proposed frameworks. For SVR, the gaussian Radial Basis Function (RBF) is used as kernel function, kernel coefficient γ is set to 0.1, the bandwidth of the RBF kernel ϵ is set to 1.4, and penalty parameter C is set to 500. For GBRT, the number of trees is 500, the learning rate ν is 0.1, and the Least Absolute Deviation (LAD) is used to optimize the loss function. For the MLP regressor, the number of hidden layers is 1 with 50 hidden units, activation function for the hidden layer is Rectified Linear Unit (ReLU) function, and the solver for weight optimization is Limited-memory Broyden Fletcher Goldfarb Shanno (L-BFGS) method. For both LSTM and GRU, four hidden layers each of which has fifty hidden units are stacked together. Other hyperparameters such as activation function, dropout rate, and epoch are set to ReLU, 0.2 and 50 respectively. Table 1 presents all the necessary parameters of the forecasters. The choice of the above-mentioned parameters is based on the trial and error method.

C. ACCURACY IMPROVEMENT CAPABILITY OF A-SLET

To study the effectiveness and stability of A-SLET, experiments are carried out in four different cases. In the first two cases, A-SLET is applied to forecast the daily peak load of the ME load zone of ISO-NE for the years 2014-2015 and 2015-2016. For next two cases, experiments are carried out to forecast the daily peak load of NSW for the years 2014-2015 and 2015-2016. All the cases are considered from the beginning of March to the end of February the following year. Experiment to study the effectiveness of the A-SLET is carried out on season basis because peak load pattern varies from season to season and seasonal experiments can explore the effectiveness of the strategy regardless of load pattern. Fig. 4 shows the peak load variation of different seasons. The order of the seasons in ME and NSW are spring, summer, fall, and winter. Each season has three months of duration, and for the ME region, spring starts from the beginning of March and follows the order. For the NSW region, spring starts from September and follows the order. It is evident from Fig. 4 that there are some similarities and dissimilarities in

TABLE 1. Parameters and hyperparameters of the forecasters.

Forecaster	Parameters and hyperparameters	Value
SVR	Kernel function	RBF
	Kernel coefficient	0.1
	Bandwidth of the RBF kernel	1.4
	Penalty parameter	500
GBRT	Number of trees	500
	Learning rate	0.1
	Loss function optimizer	LAD
MLP	Number of hidden layers	1
	Hidden units	50
	Activation function	ReLU
	Weight optimization solver	L-BFGS
LSTM	Total number of hidden layers	4
	Hidden unit in each hidden layer	50
	Activation function	ReLU
	Dropout rate	0.2
	Number of epochs	50
GRU	Size of mini-batch	10
	Total number of hidden layers	4
	Hidden unit in each hidden layer	50
	Activation function	ReLU
	Dropout rate	0.2
	Number of epochs	50
	Size of mini-batch	10

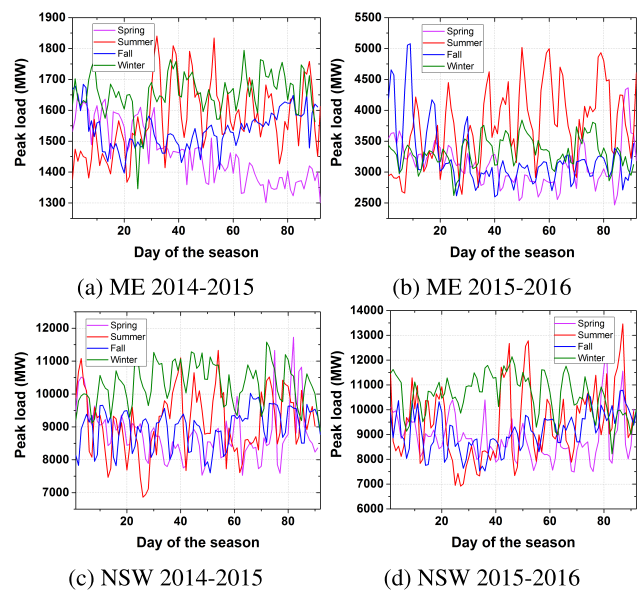


FIGURE 4. Peak load of four seasons for different cases.

the load pattern of two different load zones. Four years of historical data before the forecasted horizon are used as the training set for final forecasting. For error measurement and error trend calculation, a few more years of historical data are used.

In this experiment, error measurement is done by ensemble forecaster discussed in section II.C. Six-degree polynomial interpolation is used to get the error trend from the previous

TABLE 2. Seasonal MAPE of different forecasting techniques.

Method	ME 2014-2015				ME 2015-2016				NSW 2014-2015				NSW 2015-2016			
	Fall.	Win.	Spr.	Sum.	Fall.	Win.	Spr.	Sum.	Fall.	Win.	Spr.	Sum.	Fall.	Win.	Spr.	Sum.
SVR	2.15	2.37	2.89	2.31	2.33	2.38	2.41	2.24	2.36	2.16	3.98	5.69	3.30	2.92	4.37	8.26
SVR+A-SLET	1.60	1.87	2.36	2.17	1.89	2.20	1.87	2.13	2.10	2.06	3.42	4.96	3.00	2.54	4.33	7.83
MLP	2.24	2.26	2.51	2.99	1.86	2.23	3.06	2.05	2.83	2.02	3.08	4.33	2.71	2.20	3.73	6.92
MLP+A-SLET	1.76	1.76	1.80	3.46	1.70	1.71	2.51	1.84	2.61	1.66	2.76	4.20	2.44	1.79	3.43	4.65
GBRT	1.70	2.37	2.53	4.26	3.00	2.30	2.37	2.55	2.35	2.66	3.23	5.25	2.94	2.73	3.31	4.81
GBRT+A-SLET	1.91	1.84	2.88	3.49	2.07	2.13	1.94	2.10	1.92	2.03	2.98	4.87	2.65	2.44	3.21	4.70
LSTM	3.80	3.21	5.41	5.53	3.31	2.57	4.55	2.94	2.95	2.27	5.07	5.61	4.16	3.48	5.81	6.63
LSTM+A-SLET	2.59	2.40	3.82	3.53	2.84	2.18	2.87	2.54	2.47	2.13	3.56	4.63	2.75	3.65	3.59	6.31
GRU	2.50	3.77	5.62	2.68	3.57	2.65	3.97	3.72	3.28	2.26	5.18	5.80	4.41	3.94	4.85	6.39
GRU+A-SLET	2.47	2.13	2.30	4.08	1.91	2.57	3.20	3.59	2.90	1.86	4.51	4.27	2.84	2.69	4.37	5.84

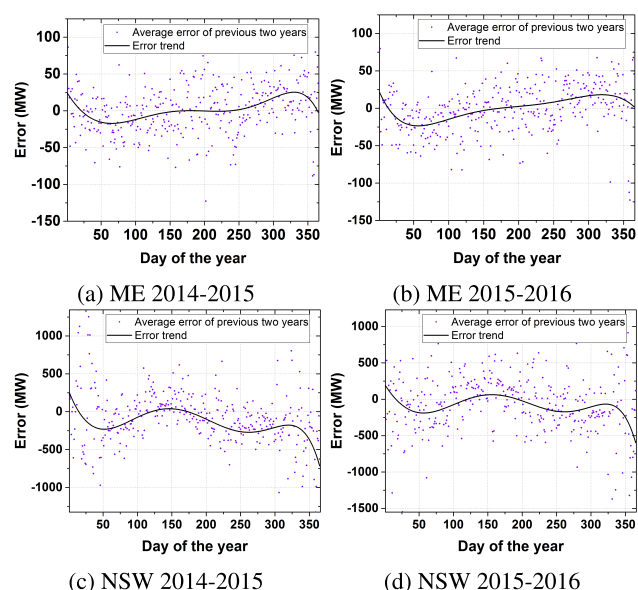


FIGURE 5. Error trend of different cases.

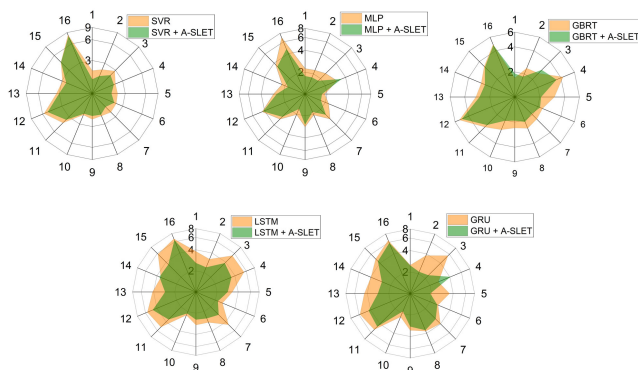


FIGURE 6. Seasonal MAPE of different forecasting techniques. Axis 1 to 16 represents the 16 seasons as in the order of left to right of Table 2.

two year’s average error. Fig. 5 shows the error trend of four cases. It is quite evident from the figure that, error trend of the same region follows almost similar error patterns.

Table 2 and Fig. 6 show forecasting errors for different forecasting methods. In the experiment, SVR, GBRT, and

TABLE 3. Standard deviation of peak load for different seasons.

Case	Spring	Summer	Fall	Winter
ME 2014-2015	94.08	115.42	63.54	69.28
ME 2015-2016	99.63	123.92	89.65	80.31
NSW 2014-2015	808.95	981.47	586.4	638.41
NSW 2015-2016	933.92	1487.63	789.93	814.86

MLP separately served as adaptive forecaster of A-SLET. Results demonstrate the inclusion of our technique for SVR, GBRT, and MLP increases the accuracy in almost all scenarios which indicate the proposed method is forecaster independent. In other words, A-SLET increases the forecasting accuracy regardless of the forecaster that is used.

Furthermore, Table 2 and Fig. 6 show seasonal forecasting errors. As each year has four seasons, the three forecasters have 20 different scenarios. Table 2 and Fig. 6 present 80 seasonal scenarios as four cases are considered for the analysis. In most scenarios, our proposed forecasting technique works well. The higher the standard deviation is for any season the more the load fluctuates in that season. A-SLET works well even when the load fluctuates excessively in summer (See Table 3). Although many features have been considered including meteorological, calendar, and economic features to track the peak load of ME and NSW, forecasting error for base forecasting method is higher than A-SLET. The reason behind it is the fact of unknown features that were not considered. Aiming to solve this challenge, the proposed A-SLET calculates error trend of historical load to reveal the trend effect caused by these unknown load affecting factors. Then, the second learning of error trend in A-SLET gains the underlying laws of historical error trend effect for the future load. Thus, A-SLET transforms the unknown load affecting factors into known forecasting features and improves the accuracy of load forecasting. As a result, error reduction of summer for ME 2014-2015 (GBRT), NSW 2014-2015 (SVR), NSW 2015-2016 (MLP), ME 2014-2015 (LSTM),

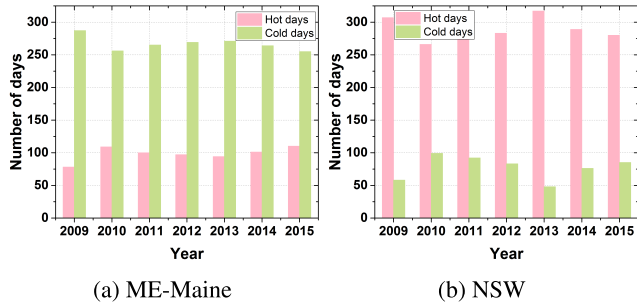


FIGURE 7. Distribution of hot days and cold days of ME-Maine and NSW.

and NSW 2014-2015 (GRU) are 18.07%, 12.83%, 32.80%, 36.16%, and 26.27% respectively, and average error reduction for all 80 scenarios is 15.17% in terms of MAPE.

Out of the 80 scenarios, A-SLET is incapable of just five scenarios (ME 2014-2015-Sum.-MLP, ME 2014-2015-Fall-GBRT, ME 2014-2015-Spr.-GBRT, NSW 2015-2016-Win.-LSTM, and ME 2014-2015-Sum.-GRU) and the best performing scenarios are ME 2014-2015-Fall-SVR, NSW 2015-2016-Sum.-MLP, NSW 2014-2015-Win.-GBRT, NSW 2015-2016-Spr.-LSTM, and ME 2014-2015-Spr.-GRU. The incapable scenarios and best performing scenarios are distributed in different geographical areas and seasons. Our proposed method did not perform well on these five scenarios because of the load structure of the specific forecasted horizon. In other words, our proposed technique is general to almost all scenarios regardless of seasonal and geographical differences. Furthermore, A-SLET can improve the accuracy of all the yearly scenarios. As the cases are from two vast geographical areas with distinct load patterns, seasonal and yearly analysis of the cases show the generality of the method.

D. ACCURACY IMPROVEMENT CAPABILITY OF AP-SLET

To demonstrate the effectiveness of AP-SLET, experiments are carried out with the same cases mentioned earlier. A yearly forecasting horizon is considered to analyze the effectiveness of the proposed technique. The climate of the two specified regions is different from each other. ME is in the hemiboreal climate region and NSW in the humid subtropical climate region. For balanced distribution of cold days and hot days in the training set, hot days in the training set can be used to train the adaptive forecaster for hot days and cold days in the same training set can be used for training the adaptive forecaster for cold days. However, for imbalance distribution, domain knowledge can be applied for selecting the right length of the training set. The distribution of hot days and cold days for ME and NSW are shown in Fig. 7. It is evident from the figure that ME has more cold days than the hot days and vice versa for NSW. For the proper training of the forecaster, because of the imbalanced distribution of hot days and cold days, the whole training set is used to train the adaptive forecaster for hot days in case of ME and whole training set to train the adaptive forecaster for cold days in case of NSW.

TABLE 4. Yearly MAPE of different forecasting techniques.

Method	ME 2014-2015	ME 2015-2016	NSW 2014-2015	NSW 2015-2016
SVR	2.44	2.44	3.54	4.74
SVR+A-SLET	2.00	2.04	3.13	4.45
SVR+AP-SLET	1.89	2.07	3.13	4.41
MLP	2.42	2.77	3.35	3.47
MLP+A-SLET	2.32	2.58	3.04	3.30
MLP+AP-SLET	2.24	2.25	2.81	3.14
GBRT	2.47	2.11	3.40	3.37
GBRT+A-SLET	2.14	1.98	3.07	3.19
GBRT+AP-SLET	2.10	1.87	3.05	3.18
LSTM	4.48	5.99	3.68	4.63
LSTM+A-SLET	2.44	2.86	3.21	4.41
LSTM+AP-SLET	2.24	4.04	3.16	4.18
GRU	4.12	4.49	3.16	5.11
GRU+A-SLET	2.35	2.31	3.01	4.33
GRU+AP-SLET	2.19	2.96	2.98	3.86

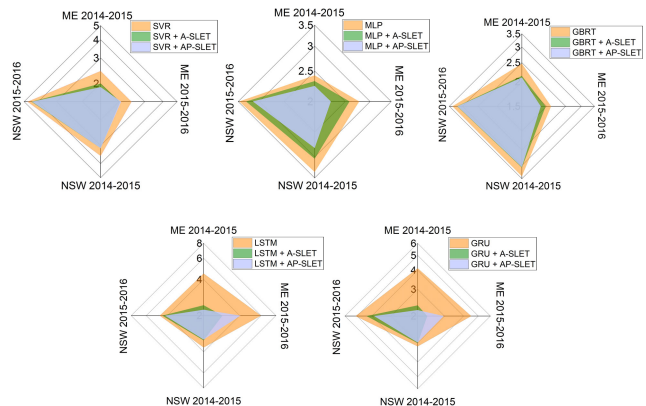


FIGURE 8. Yearly MAPE of different forecasting techniques.

Similar to the A-SLET experiment, SVR, GBRT, MLP, LSTM, and GRU separately serve as adaptive forecaster of AP-SLET and compared their performance with corresponding A-SLET, and base forecaster. The experiments are carried out in twenty different scenarios. Each scenario is comprised of a forecaster, forecaster with A-SLET, and AP-SLET. The experiment results are given in Table 4 and Fig. 8. Analyzing the table and figure, it is found that out of the twenty scenarios, the AP-SLET does not perform well in three scenarios and in one scenario MAPE is equal compared with A-SLET. However, AP-SLET demonstrates better forecasting accuracy in all the scenarios compared with base forecasters. MAPE of AP-SLET is reduced by 18.39% while MAPE of A-SLET is reduced by 16.93% on average. MAPE of AP-SLET is better compared to A-SLET because AP-SLET involves training set classification method based on the balance point temperature that allows AP-SLET to learn better than A-SLET. However,

TABLE 5. MAPE comparison with [30].

Method	ME (Jan 2017 - Feb 2017) MAPE
[30]	1.82
AP-SLET	1.51

TABLE 6. MAPE comparison with [9].

Method	ISO-NE (2008-2009)	NSW (Jan.-Mar 2017)	ERCOT (2015-2016)
ANN-FTL [9]	3.21	6.05	5.28
AP-SLET	2.49	4.97	3.52

AP-SLET shows slightly better performance than A-SLET because training set classification and parallel forecasting have less impact on accuracy improvement compared to adaptive second learning.

E. COMPARATIVE EXPERIMENT ON AP-SLET AND STATE-OF-THE-ART FORECASTING METHODS

In this section, a comparative study is conducted on AP-SLET, state of the art forecasting method [30], ANN-FTL [9], and BART [39]. Reference [30] proposed a training set construction method based on the day-to-day topological network. ANN-FTL is a hybrid forecasting method that integrates FTL with ANN while BART is a non-parametric, Bayesian, sum-of-trees model. Similar to the mentioned state-of-the-art forecasting methods, features considered for the comparison are calendar, weather, socio-economic, and historical load data. The ensemble forecaster described in section II.C used for error measurement and adaptive forecaster of AP-SLET. Both [30] and AP-SLET involve technique for proper training the forecasting model with training samples similar to the test samples. While [30] builds topological network of similar days, AP-SLET uses balance point temperature-based training set classification technique for training the model. As shown in Table 5, the MAPE of AP-SLET is 1.51% and MAPE of [30] is 1.82% for the same test case, which indicates AP-SLET shows better forecasting accuracy.

The comparison with [9] is carried on three datasets: ISO-NE [36], NSW [37], and ERCOT [40]. For all the three datasets, both ANN-FTL and AP-SLET are applied to the same test set mentioned in [9]. Analyzing Table 6, it is found that AP-SLET is the best performer for all three cases and MAPE of AP-SLET is reduced by 17.85%-33.33%.

To compare the result with BART, both BART and AP-SLET are applied to ERCOT [40]. Similar to [39], detrending is performed prior to applying AP-SLET. As shown in Table 7, compared to BART, RMSE, and MAE of AP-SLET are reduced by 34.05% and 35.38% respectively.

AP-SLET usually requires a long training set in comparison to the technique where there is no training set construction technique involved. This is because the techniques with

TABLE 7. Comparing performance metrics with [39].

Method	Out-of-sample error (ERCOT 2002-2017)	
	RMSE	MAE
BART [39]	2.866	2.213
AP-SLET	1.89	1.43

the training set construction use a part of the training set that is similar to the test set instead of the whole training set. Moreover, for imbalance distribution of hot days and cold days in the dataset requires domain knowledge to set the right length of the training set.

Time consumption of AP-SLET is compared with [30] using the same training set and test set. The overall running time of [30] and AP-SLET is 24.7419(h) and 0.0055(h) respectively, where the overall running time is the sum of learning and testing time. For our proposed method, learning time is used for error measurement, curve fitting, and second learning. The overall running time of AP-SLET is much lower compared to [30] because AP-SLET uses simple training set classification method where [30] uses a very complex day-to-day topological network. The time-cost comparison shows AP-SLET fully meets the real-life application requirement.

V. CONCLUSION

The major contribution of this paper focuses on improving the forecasting accuracy by adapting the error trend in the second learning caused by unknown load affecting factors. A new strategy is proposed in this paper called A-SLET. A robust experiment has been done to show the effectiveness of A-SLET. The A-SLET has been examined in eighty different scenarios, which are comprised of various forecasters, forecasting horizon, and areas with a vast geographical difference. Compared to the base forecaster, MAPE of A-SLET is reduced by 15.17% and 16.93% on average for seasonal and yearly forecasting respectively. To further extend, along with a proposed training set classification method and parallel forecasting, AP-SLET is proposed. It is shown from the experimental results that the accuracy of AP-SLET is better compared to the A-SLET in most of the scenarios and MAPE of AP-SLET is reduced by 18.39% on an average compared to the base forecaster. AP-SLET is further compared with three state-of-art forecasting methods. Compared to the state-of-art forecasting methods, MAPE, RMSE, and MAE of AP-SLET are reduced by 17.03%-33.33%, 34.05%, and 35.38% respectively. The experimental results demonstrate both A-SLET and AP-SLET can transform the unknown and unavailable load affecting factors into known forecasting features and then adapt it to improve forecasting performance. Both of the strategies are also equally applicable to almost all load scenarios ignoring geographical and seasonal differences.

REFERENCES

- [1] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electr. Power Syst. Res.*, vol. 142, pp. 58–73, Jan. 2017.

- [2] M. Jin, X. Zhou, Z. M. Zhang, and M. M. Tentzeris, "Short-term power load forecasting using grey correlation contest modeling," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 773–779, Jan. 2012.
- [3] T. Ahmed, D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "Load forecasting under changing climatic conditions for the city of Sydney, Australia," *Energy*, vol. 142, pp. 911–919, Jan. 2018.
- [4] G. Cerne, D. Dovzan, and I. Skrjanc, "Short-term load forecasting by separating daily profiles and using a single fuzzy model across the entire domain," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7406–7415, Sep. 2018.
- [5] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, "Long-term load forecasting for fast developing utility using a knowledge-based expert system," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 491–496, May 2002.
- [6] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1087–1088, Jan. 2018.
- [7] M. Karimi, H. Karami, M. Gholami, H. Khatibzadehazad, and N. Moslemi, "Priority index considering temperature and date proximity for selection of similar days in knowledge-based short term load forecasting method," *Energy*, vol. 144, pp. 928–940, Feb. 2018.
- [8] G. Dudek, "Artificial immune system with local feature selection for short-term load forecasting," *IEEE Trans. Evol. Comput.*, vol. 21, no. 1, pp. 116–130, Feb. 2017.
- [9] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Appl. Energy*, vol. 217, pp. 537–549, May 2018.
- [10] M. Barman, N. B. D. Choudhury, and S. Sutradhar, "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in assam, India," *Energy*, vol. 145, pp. 710–720, Feb. 2018.
- [11] W. Ma, S. Fang, G. Liu, and R. Zhou, "Modeling of district load forecasting for distributed energy system," *Appl. Energy*, vol. 204, pp. 181–205, Oct. 2017.
- [12] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [13] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang, "Short-term load forecasting for electric vehicle charging stations based on deep learning approaches," *Appl. Sci.*, vol. 9, no. 9, p. 1723, Apr. 2019.
- [14] N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu, "A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids," *Appl. Energy*, vol. 129, pp. 336–345, Sep. 2014.
- [15] J. Wang, W. Yang, P. Du, and Y. Li, "Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system," *Energy*, vol. 148, pp. 59–78, Apr. 2018.
- [16] H. Zhao and S. Guo, "An optimized grey model for annual power load forecasting," *Energy*, vol. 107, pp. 272–286, Jul. 2016.
- [17] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Appl. Energy*, vol. 167, pp. 135–153, Apr. 2016.
- [18] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, Mar. 2016.
- [19] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22–29, May 2016.
- [20] 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.
- [21] Y. Liu, W. Wang, and N. Ghadimi, "Electricity load forecasting by an improved forecast engine for building level consumers," *Energy*, vol. 139, pp. 18–30, Nov. 2017.
- [22] M. Zhou and M. Jin, "Holographic ensemble forecasting method for short-term power load," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 425–434, Jan. 2019.
- [23] Q. Xu, X. Yang, and X. Huang, "Ensemble residual networks for short-term load forecasting," *IEEE Access*, vol. 8, pp. 64750–64759, 2020.
- [24] P. Mandal, T. Senjyu, N. Urasaki, and T. Funabashi, "A neural network based several-hour-ahead electric load forecasting using similar days approach," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 6, pp. 367–373, Jul. 2006.
- [25] S. Barghinia, S. Kamankesh, N. Mahdavi, A. H. Vahabie, and A. A. Gorji, "A combination method for short term load forecasting used in iran electricity market by NeuroFuzzy, Bayesian and finding similar days methods," in *Proc. 5th Int. Conf. Eur. Electr. Market*, May 2008, p. 5.
- [26] P. Mandal, A. K. Srivastava, and J.-W. Park, "An effort to optimize similar days parameters for ANN-based electricity price forecasting," *IEEE Trans. Ind. Appl.*, vol. 45, no. 5, pp. 1888–1896, Sep./Oct. 2009.
- [27] Q. Mu, Y. Wu, X. Pan, L. Huang, and X. Li, "Short-term load forecasting using improved similar days method," in *Proc. Asia-Pacific Power Energy Eng. Conf.*, 2010, pp. 1–4.
- [28] Y. Chen, P. B. Luh, C. Guan, Y. Zhao, L. D. Michel, M. A. Coolbeth, P. B. Friedland, and S. J. Rourke, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, Feb. 2010.
- [29] P. Li, J. Zhang, C. Li, B. Zhou, Y. Zhang, M. Zhu, and N. Li, "Dynamic similar sub-series selection method for time series forecasting," *IEEE Access*, vol. 6, pp. 32532–32542, 2018.
- [30] P. Zeng and M. Jin, "Peak load forecasting based on multi-source data and day-to-day topological network," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 6, pp. 1374–1381, Mar. 2018.
- [31] P. Lusi, K. R. Khalilpour, L. Andrew, and A. Liebman, "Short-term residential load forecasting: Impact of calendar effects and forecast granularity," *Appl. Energy*, vol. 205, pp. 654–669, Nov. 2017.
- [32] B. Pavlyshenko, "Using stacking approaches for machine learning models," in *Proc. IEEE 2nd Int. Conf. Data Stream Mining Process. (DSMP)*, Aug. 2018, pp. 255–258.
- [33] X. Luo, J. Sun, L. Wang, W. Wang, W. Zhao, J. Wu, J.-H. Wang, and Z. Zhang, "Short-term wind speed forecasting via stacked extreme learning machine with generalized coreentropy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4963–4971, Nov. 2018.
- [34] 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.
- [35] N. Ghadimi, A. Akbarimajd, H. Shayeghi, and O. Abedinia, "Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting," *Energy*, vol. 161, pp. 130–142, Oct. 2018.
- [36] ISO New England Inc. *ISO-NE*. Accessed: Oct. 26, 2017. [Online]. Available: <https://www.iso-ne.com/isoexpress/web/reports/pricing/-tree/zone-info>
- [37] Australian Energy Market Operator. *AEMO*. Accessed: Oct. 26, 2017. [Online]. Available: <https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data-dashboard>
- [38] TWC Product and Technology LLC 2014–2020. *Weather Underground*. Accessed: Oct. 26, 2017. [Online]. Available: <http://api.wunderground.com/history/>
- [39] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the Texas region," *Energy*, vol. 185, pp. 1143–1153, Oct. 2019.
- [40] Electric Reliability Council of Texas. *ERCOT*. Accessed: Oct. 26, 2017. [Online]. Available: http://www.ercot.com/gridinfo/load/load_hist

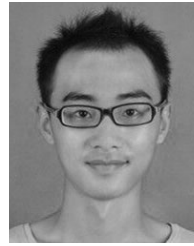


MD FAZLA ELAHE was born in Bogura, Bangladesh, in 1991. He received the B.S. degree in software engineering from the Shenyang University of Chemical Technology, Shenyang, China, in 2012, and the M.S. degree in computer software and theory from the Nanjing University of Science and Technology, Nanjing, China, in 2017. He is currently pursuing the Ph.D. degree in computer science and technology with Hunan University, Changsha, China. Since 2013, he has been a Lecturer with the Department of Software Engineering, Daffodil International University, Bangladesh. His research interests include data mining, artificial intelligence, power load forecasting, and bioinformatics.



MIN JIN (Member, IEEE) received the B.E. degree in automation and the M.S. degree in control theory and engineering from the Central South University of Technology, Changsha, China, in 1995 and 1997, respectively, and the Ph.D. degree in control theory and engineering from Central South University, Changsha, in 2000.

From August 2000 to August 2004, she was a Senior Engineer with Dongfang Electronics Information Industry Company, Yantai, China. From December 2010 to January 2012, she was a Visiting Scholar with the Department of Electrical and Computer Science, Georgia Institute of Technology, Atlanta, GA, USA. Since November 2004, she has been with the College of Computer Science and Electronic Engineering, Hunan University, where she is currently a Professor. Her research interests include artificial intelligence, industrial big data, embedded systems, and their engineering applications.



PAN ZENG (Associate Member, IEEE) received the B.E. degree in building electricity and intelligence from the Qingdao University of Technology, Qingdao, China, in 2012, and the M.S. degree in software engineering from Hunan University, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering. His research interests include machine learning, artificial intelligence, load forecasting, and their engineering applications.

• • •