



Research article

EMLP: short-term gas load forecasting based on ensemble multilayer perceptron with adaptive weight correction

Fengyong Li*, Meng Sun

College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 201306, China

* **Correspondence:** Email: fyli@shiep.edu.cn.

Abstract: This paper tackles a recent challenge in smart city that how to improve the accuracy of short-term natural gas load forecasting. Existing works on natural gas forecasting mostly reply on a combined forecasting model by simply integrating several single-forecasting models. However, due to the existence of redundant single-forecasting models, these works may not attain a higher prediction accuracy. To address the problem, we design a new natural gas load forecasting scheme based on ensemble multilayer perceptron (EMLP) with adaptive weight correction. Our method firstly normalizes multi-source data as original data set, which is further segmented by a window model. Then, the abnormal data is removed and subsequently interpolated to form a complete normalized data set. Furthermore, we integrate a series of multilayer perceptron (MLP) network to construct an ensemble forecasting model. An adaptive weight correction function is introduced to dynamically modify the weight of the previous predicted result. Since the correction function can match well the volatility characteristics of load data, the prediction accuracy is significantly improved. Extensive experiments demonstrate that our method outperforms existing state-of-the-art load forecasting schemes in terms of the prediction accuracy and stability.

Keywords: load forecasting; ensemble learning; multilayer perceptron; smart city

1. Introduction

With the rapid development of Internet of things, big data, cloud computing and the energy revolution [1], the concept of “smart gas” has emerged as an important part of smart city. Smart gas is oriented by the construction of intelligent pipe networks and user experience, and finally realize the intelligence of gas network. Accordingly, how to ensure the reliability of natural gas supply has become an urgent issue in smart city [2]. Since the reliable supply depends on the accurate prediction of gas load, load forecasting plays a critical role in the maintenance of smart natural gas development.

The gas load is a dynamic system and may be changed with different regions and different times. Accordingly, it is very difficult for forecasters to choose an appropriate forecasting model that can match well the law of load development in this region. It is still an open problem to accurately select a prediction model suitable for local conditions. A number of researchers have reported many works. In these solutions, traditional mathematical algorithms and statistical prediction methods, e.g., regression analysis [3, 4], are employed to address the above challenges. Nevertheless, economic development and different user habits always lead to a large fluctuation in the total gas load. Accordingly, traditional regression methods cannot adapt to the non-linearity variability of the current gas daily load data [5], resulting in that the forecast results are far from the actual gas consume. With the rise of deep learning methods, some researchers use deep neural network to build gas load forecasting models [6–8], which can overcome the problem that traditional schemes converge slowly and easily fall into local minima. However, these solutions rarely consider complexity requirements when they are designed, leading to a high time-consuming [8]. Another research is the combined forecasting [9–11] that employs two or more single load forecasting methods to build a integrated model, and then select the linear or nonlinear weight coefficient to realize combined forecasting. Combined forecasting model, however, is also difficult to give a satisfactory solution due to the following two pitfalls. First, some redundant models maybe appear when combined forecasting involves a large number of single models. This leads to a decrease in prediction accuracy. Second, the forecasting results of the same model in different time periods are significantly different. If the weights of different single models in combined forecasting model always keep the same values, the forecast results cannot meet completely with the actual situation.

One might think that traditional mature load forecasting methods, e.g., power load forecasting [12–15], can also solve the above problem. Nevertheless, power load forecasting technology is hard, at least inconvenient, to directly apply in our context. This is because, compared with power load data, gas load data has a wider range of fluctuations so that the prediction process prone to over-fitting. Although some power load forecasting works use packet loss technology to solve the over-fitting problem [16, 17], it further leads to the destruction of the integrity of data set and finally affects the prediction results.

Obviously, gas load forecasting cannot directly apply existing industry forecasting methods, it may require some special processing. We are thus motivated to design a new gas load forecasting scheme based on ensemble multilayer perceptron with adaptive weight correction. Proposed scheme can not only provide a higher prediction accuracy for short-term gas load, but also effectively avoid the over-fitting problem in training procedure.

In general, the contributions of the paper are as follows:

- We design a new gas load forecasting scheme based on ensemble learning, which can not only achieve high performance in short-term load forecasting, but can effectively avoid the over-fitting problem caused by gas load fluctuations in training procedure.
- Our method can work over multiple gas equipments. The abnormal values in original data set are firstly cleaned by designing a window function to form a complete normalized data set. This makes the original data set can be forecasted correctly. Subsequently, a series of multilayer perceptron networks are integrated to construct an ensemble forecasting model, in which a correction function is further introduced to dynamically modify the weight of the prediction result. Since the introduced correction function can match well the volatility and increasing characteristics of daily gas load data, the prediction accuracy is significantly improved.

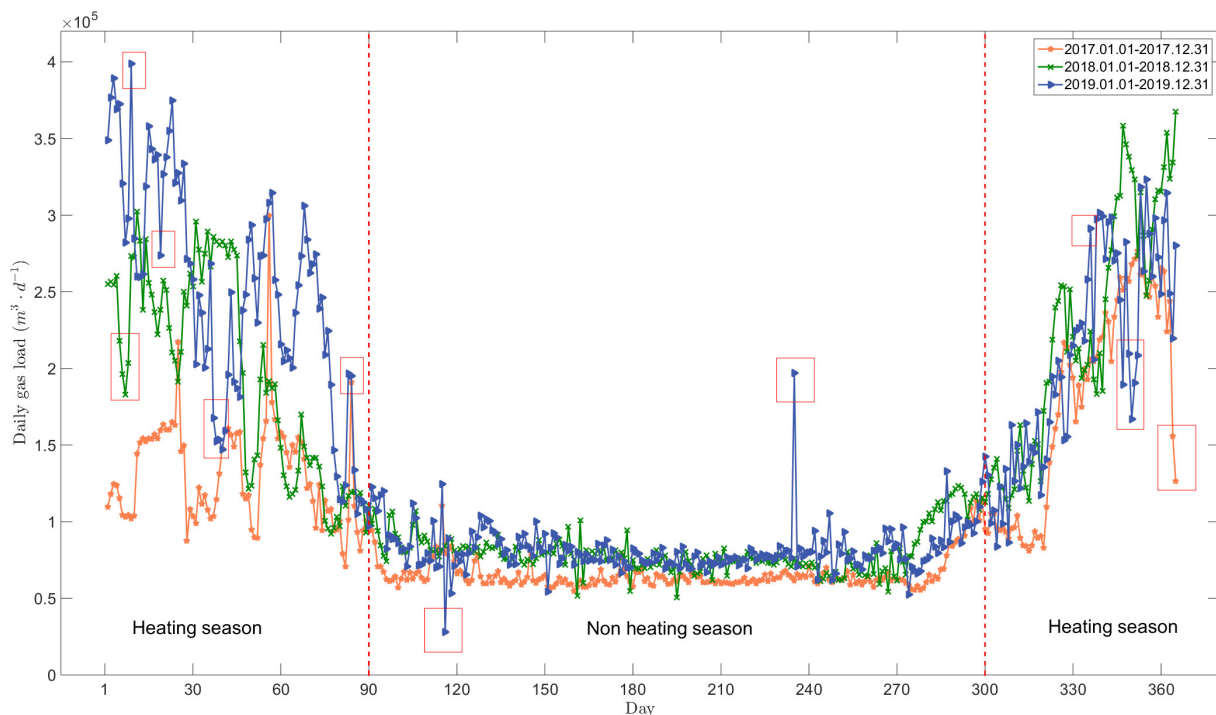


Figure 1. The daily gas load data of a city of southern China from 2017 to 2019. The red boxes indicate the part of abnormal load data in heating season and non-heating season.

- We perform comprehensive experiments by comparing with multiple well-known learning methods. Experimental results show that our scheme can accurately forecast the daily gas load and outperforms the state-of-the-art in terms of the prediction accuracy.

The rest of this paper is organized as follows. Section 2 introduces the previous gas load forecasting works. The detailed procedure of proposed scheme are shown in Section 4. Subsequently, comprehensive experiments are performed to evaluate the performance of proposed scheme. The experimental results and corresponding discussions are presented in Sections 5. Finally, Section 6 concludes the paper.

2. Related works

Existing combined forecasting solutions mainly includes two categories: horizontal combination forecasting model and vertical combination forecasting model. The former uses two or more load forecasting methods to predict separately, and then introduce linear or nonlinear weight coefficient to achieve combined forecasting, while the latter mainly employs the results of one or more prediction methods to guide the parameter selection or result correction in other prediction methods. We briefly introduce existing works according to these two categories.

Regarding the horizontal combination forecasting model, several state-of-the-art forecasting schemes have been developed [9–11, 18]. Ervural et al. [9] combined autoregressive moving average method and genetic algorithm to construct model and implement daily gas load forecasting. This

combined model has a strong robust and better than any single model in terms of average relative error and cost function value. Panapakidis et al. [10] tested the robustness of a novel hybrid computational intelligence predictions model by combining the Wavelet Transform (WT), Genetic Algorithm (GA), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Feed-Forward Neural Network (FFNN), and finally obtained a combination prediction model with better prediction effect. Qiao et al. [11] designed a hybrid prediction model that integrates an improved whale swarm algorithm (IWOA) and relevance vector machine (RVM). Proposed scheme has a higher prediction accuracy when the amount of data is larger or smaller, but, the calculation time is relatively long. Yu and Xu et al. [18] proposed an appropriate combinational approach which is based on improved BP neural network for short-term gas load forecasting, and the network is optimized by the real-coded genetic algorithm. As a result, the integration model improved by modified additional momentum factor gets more ideal solutions for short-term gas load forecasting.

Regarding the vertical combination forecasting model, a considerable literature has grown up around this theme [19–21]. Ulrich et al. [19] improved the kernel function of support vector machine (SVM) and grid search of MLP networks by wavelet analysis, and then get a significant improvement of forecasting performance. Taspinar et al. [20] employed the residual value sequence calculated by the gray theory model and the output vector obtained by the fuzzy theory to build the input vector of recurrent neural network model. The forecasting performance is superior to original recurrent neural network model. Zu et al. [21] used autoregressive integrated moving average model (ARIMA) and BP neural network to perform load forecasting, and then employed information entropy theory to weight the results.

According to the above discussion, combined forecasting methods have been developed to address the gas load forecasting problem. Many existing methods have made great efforts to improve the forecasting accuracy by introducing existing data processing technology, such as regression analysis, wavelet analysis and neural network. Nevertheless, these existing methods still have several obvious disadvantages: (1) The over-fitting problem for small-scale load data set still exists. (2) The generalization performance of existing prediction model is relatively weak. (3) The prediction accuracy needs to be improved further. This makes that they cannot match well the actual characteristics of gas load data, and thus may have a weak forecasting ability. Some works may simply borrow the idea from the power load forecasting methods [10, 18], but, different from power load data, gas load data has a wider range of fluctuations, the forecasting process can easily fall into an over-fitting state and thus do not perform well. To the best of our knowledge, too little gas load forecasting work can offer a satisfactory solution for the above disadvantages. This paper tries to fill these gaps.

3. Motivating observation

The gas load data is considered as the total gas consumption of all users in the gas company, whose unit is usually defined as $(m^3 \cdot d^{-1})$. Since natural gas is usually used for urban heating, the load data is greatly affected by temperature, that is, when the temperature changes significantly, the daily load data will generally change accordingly.

The following experiments can verify the relationship between temperature and the daily gas load data. Figure 1 shows the actual daily gas load data of a city of southern China from 2017 to 2019. The

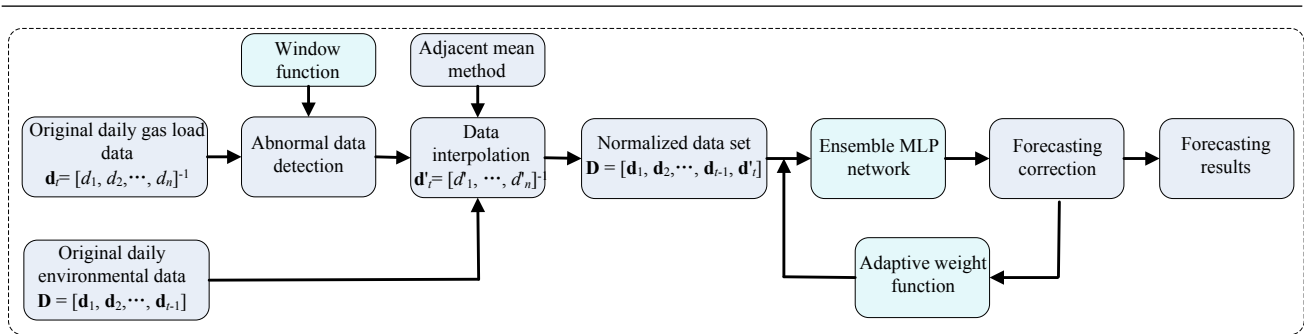


Figure 2. The framework of proposed scheme.

load data contains daily consumption data for 365 days of each year*. From this figure, we can obtain three observations as follows. **First**, the daily gas load data is obvious higher in spring and winter (called as *heating season*), while lower in summer and autumn (called as *non-heating season*). This verifies the conclusion that the daily gas load data is greatly affected by the temperature. **Second**, with the increase of the year, the daily gas load data at the peak of gas consumption also increase significantly. This implies that the gas consumption shows a trend of increasing year by year. Therefore, in order to accurately predict the gas load data, the forecasting model needs to be adjusted in real time based on short-term prior data. **Third**, we can see that there are a lot of abnormal data (marked by red boxes in Figure 1) in daily gas load data, especially in heating season. For most forecasting models, since the prediction accurate mainly depends on prior data, we can implement data cleaning before model training to further improve the prediction performance.

4. Proposed scheme

4.1. Framework of proposed scheme

In this section, we design an ensemble prediction method to improve the precision of short-term gas load forecasting. The detailed framework of proposed scheme is shown in Figure 2. According to this framework, proposed scheme is mainly comprised of two parts. In the former, multi-source data is input as original data set, which is then segmented by a window model to detect the abnormal data. Then, the abnormal values are removed and further interpolated by adjacent mean to form a complete normalized data set. In the latter, an ensemble prediction model is constructed by integrating a series of single MLP network. Each forecasting result from single MLP network is dynamically corrected by an adaptive weight, which is calculated from the prior short-term data. Finally, the corrected data is output as the final forecasting results.

4.2. Data processing based on window model

4.2.1. Data pre-processing

Since the daily gas load data generally changes with the temperature, the maximum temperature, minimum temperature and average temperature are considered to be the top three factors affecting the daily load data. In addition, different dates and weather conditions also cause fluctuations for the

*In general, load data is sampled every half an hour and thus produce 48 values each day, which are aggregated as daily load.

Table 1. Normalized values for seven weather conditions.

Weather type	Normalized value
Sunny	0.40
Partly cloudy	0.50
Cloudy day	0.60
Light rain	0.70
Heavy rain	0.80
Light snow	0.90
Heavy snow	1.00

daily gas load data. Therefore, a complete daily gas load feature contains six attributes: maximum temperature ($^{\circ}\text{C}$), minimum temperature ($^{\circ}\text{C}$), average temperature ($^{\circ}\text{C}$), date (M/D/Y), weather and daily load data, where the weather is limited to seven types and is normalized to the range 0 to 1 to give an ease evaluation[†]. The actual attribute values of weather are shown in Table 1.

Data pre-processing mainly contains the data integrity testing and the abnormal data cleaning, e.g., the data with its value less than 0. Assume that daily gas load data vector is represented as $\mathbf{d}_o = [d_1, d_2, \dots, d_i, \dots, d_k]$, where d_i is the i -th gas data, n is the total number of gas data. Data pre-processing procedure can be implemented easily by deleting the data with a value less than 0 and marking them as vacant data. Finally, the cleaned data will be further processed by the window model.

4.2.2. Window function construction

Daily load data always has a significant characteristic, that is, the overall fluctuation range is large, while the adjacent fluctuation range is small. Thus, we consider to construct a window model over the pre-processed data and employ this model to further normalize the original daily gas data.

Without loss of generality, assume that the pre-processed data vector is $\mathbf{d} = [d_1, d_2, \dots, d_n]$, $n \leq k$. We construct window model $\mathbf{w}_i = [d_i, d_{i+1}, \dots, d_{i+m-1}]$, where m is the width of window and $1 \leq i \leq n - m + 1$. Furthermore, the load data vector \mathbf{d} is traversed by moving the window \mathbf{w}_i . During the moving procedure, load value d_i is sequentially decided and marked by the following equation.

$$\mathbf{b}_i = \begin{cases} 0 & , \text{ if } \frac{|\bar{w} - d_i|}{\bar{w}} > E \\ 1 & , \text{ if } \frac{|\bar{w} - d_i|}{\bar{w}} \leq E \end{cases} \quad (4.1)$$

where E is the fluctuation deviation and \bar{w} is the average value of current window. \mathbf{b}_i is the state vector of current window and $b_i = 1$ represents that d_i is normal data and the abnormal data, otherwise.

$$\bar{w} = \frac{\sum_{x=i}^{i+m-1} d_x}{m} \quad (4.2)$$

[†]We investigate the actual daily gas load data of a city of southern China from 2017 to 2019. When the weather conditions are set to different parameters, they may give different prediction results. From a comprehensive comparison, when the parameters are set to the values in Table 1, the prediction model can give a more accurate result. Therefore, we apply these parameters as empirical values to the proposed model.

Accordingly, we mark the d_i as NULL if $b_i = 0$, and denote the data vector processed by window model as $\mathbf{d}' = [d'_1, d'_2, \dots, d'_n]$. Then, the adjacent mean interpolation method is used to complete \mathbf{d}' .

$$d''_i = \begin{cases} \frac{d'_{i-1} + d'_{i+1}}{2} & , \text{ if } d'_i = \text{NULL} \\ d'_i & , \text{ Otherwise} \end{cases} \quad (4.3)$$

Finally, the complete data vector after interpolation is denoted as $\mathbf{d}'' = [d''_1, d''_2, \dots, d''_n]$. Notably, since d'_{i-1} and d'_{i+1} are the two adjacent data of d'_i , respectively, they might be also NULL. If that, we move d'_{i-1} to the left or d'_{i+1} to the right until they are not NULL. In addition, if the boundary data, e.g., d'_1 and d'_n , is NULL, we consider to directly replace this data with the copy of its neighborhoods, because these data in boundary is rare, the bias is slight.

4.3. Ensemble multilayer perceptron model

4.3.1. Multilayer perceptron

Multilayer perceptron (MLP) [22] is one of the most commonly used artificial neural network algorithms. Single MLP employs bootstrap method to sample training data and then obtains multiple data subsets, which are used to train multiple sub-neural networks. Assume that the number of neurons in one sub-neural network is K , the activation function of the j -th neuron in the l -th layer can be defined as

$$net_j^l = \sum_{i=1}^n \{W_j^l D_i + b_j\} \quad (4.4)$$

where D_i is the input vector, W is the weight, b is the offset and n is the number of input vectors. In addition, define λ as the scaling factor, the suitable transfer function, e.g., sigmoid function, can be also chose to make the sub-neural network converge quickly.

$$y_j^l = \frac{1}{1 + e^{-\lambda net_j}} \quad (4.5)$$

Finally, each multilayer perceptron network (MLP_{net}) can be trained according to the aim that the mean square error (MSE) is minimized.

$$MLP_{net} = \arg \min (MSE_{(p)}) \quad (4.6)$$

where

$$MSE_{(p)} = \frac{1}{K} \sum_{j=1}^K \left(\frac{1}{n} \sum_{i=1}^{n(l)} (e_j^l)^2 \right) \quad (4.7)$$

4.3.2. Ensemble model training and correction

In this section, we integrate multiple individual multilayer perceptron networks to construct an ensemble learning model, which can be optimized by adaptively correcting the weight of forecasting result.

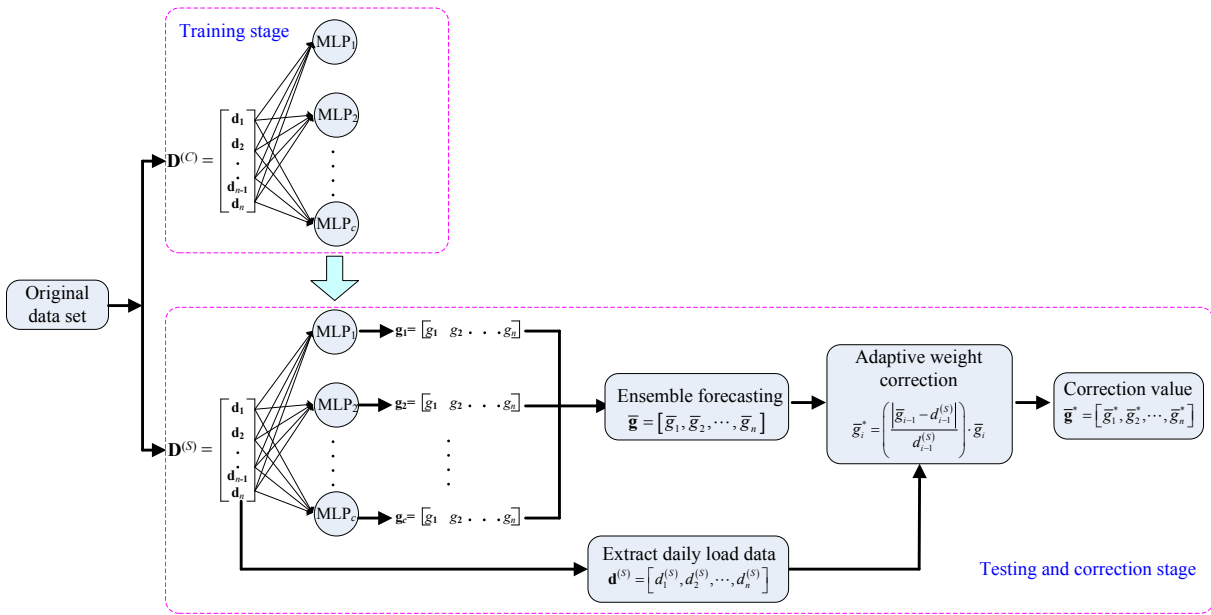


Figure 3. Training and forecasting correction procedure for proposed ensemble model.

Given a series of samples, we firstly divide these sample evenly as two parts, including training set $\mathbf{D}^{(C)} = [\mathbf{d}_1^{(C)}, \mathbf{d}_2^{(C)}, \dots, \mathbf{d}_n^{(C)}]^{-1}$ and testing set $\mathbf{D}^{(S)} = [\mathbf{d}_1^{(S)}, \mathbf{d}_2^{(S)}, \dots, \mathbf{d}_n^{(S)}]^{-1}$, where n is the number of sample in each set. $\mathbf{d}_n^{(C)}$ and $\mathbf{d}_n^{(S)}$ are the processed data vector (corresponding to \mathbf{d}'' in Section 4.2.2). The training and correcting procedures, shown in Figure 3, are also described as follows.

Step 1 : According to Equations (4.6) and (4.7), the training set $\mathbf{D}^{(C)}$ is introduced to train c MLP_{net} sequentially.

Step 2 : Combining the trained c MLP_{net} , the testing set is input to generate the forecasting result set $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_i, \dots, \mathbf{g}_n]^{-1}$, where $\mathbf{g}_i = [g_i^1, g_i^2, \dots, g_i^c]$.

Step 3 : Calculate the ensemble result $\bar{\mathbf{G}} = [\bar{g}_1, \bar{g}_2, \dots, \bar{g}_i, \dots, \bar{g}_n]$ by averaging c forecasting values,

$$\bar{g}_i = \frac{1}{c} \sum_{j=1}^c g_i^j \tag{4.8}$$

Step 4 : Sequentially correct the ensemble result $\bar{\mathbf{G}}$ by an adaptive weight α , which is calculated through the difference between the actual value and the predicted value of the previous data. Finally, the ensemble forecasting results can be further corrected as $\bar{\mathbf{G}}^* = [\bar{g}_1^*, \bar{g}_2^*, \dots, \bar{g}_i^*, \dots, \bar{g}_n^*]$.

$$\bar{g}_i^* = \left(\frac{|\bar{g}_{i-1} - d_{i-1}^{(S)}|}{d_{i-1}^{(S)}} \right) \cdot \bar{g}_i \tag{4.9}$$

Remark 1: It should be noted that for proposed ensemble model, the number of layers of each MLP network is slightly changed to generate much diversity, which is the key factor for ensemble learning [13, 24, 30]. According to Eqs (4.6) and (4.7), the training set is used to train c single MLP networks. Since single MLP network has the characteristic by itself, therefore, for the same testing data, they may give prediction results with large differences. As such, proposed scheme calculates

the ensemble results by averaging c prediction values, which can make the predicted value as close as possible to the real result.

Remark 2: In the stage of adaptive correction, since the weight is calculated from the previous testing data, the input order of the samples cannot be thus disrupted for the testing set. In other words, only after $\mathbf{d}_i^{(S)}$ is predicted, $\mathbf{d}_{i+1}^{(S)}$ can be further predicted.

5. Experimental results and discussion

In this section, a series of experiments are carried out to evaluate the effectiveness of proposed scheme.

We implement these experiments over a large-scale natural data set, which is actually collected from a city of southern China. This data set is sampled every half an hour from 2017 to 2019, so there are 51,560 data samples. In order to obtain the overall forecasting performance, we randomly select the samples from this data set each experiment to show the universality of proposed scheme. All the experiments are performed over a Windows 10 computer with an AMD(R) R7-3700x @4.2GHz, 16 GB RAM and the platform is Pycharm 2020.

Furthermore, in order to demonstrate the performance of different prediction methods, we define the average prediction error rate (*APEE*) to measure the prediction capability. Denote the original data set as $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ and the predicted data set as $\mathbf{x}' = \{x'_1, x'_2, \dots, x'_n\}$, the average prediction error can be calculated as follows. In general, for each prediction method, the lower *APEE* value means a higher prediction capability.

$$APEE = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x} - \mathbf{x}'\|^2 \times 100\% \quad (5.1)$$

In addition, we also other two evaluation metrics, Mean Absolute Error (*MAE*) and Root-Mean-Square Error (*RMSE*), to give a sufficient comparison. *MAE* mainly represents the average value of the absolute value of the error between the prediction value and the original value, while *RMSE* stands for the square root of the ratio of the sum of squares of deviations between the prediction value and the original value to the number of observations.

$$MAE = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x} - \mathbf{x}'\| \quad (5.2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}')^2} \quad (5.3)$$

5.1. Test for window model

In our scheme, the original data set are pre-processed with a window model. In order to show the advantage of the window model, we firstly test the preprocessing capability by comparing proposed scheme with two existing pre-processing models, K-means model [23] and Box model [29]. We randomly construct three sample subsets, S_1, S_2, S_3 , from the total sample set, which contain 38, 33 and 40 abnormal values, respectively.

Table 2. The accuracy of detecting abnormal samples of three data preprocessing models, Window model, K-means and Box model. Three testing results, total number of detecting abnormal samples (Total), false positive number (FP) and false negative number (FN), are shown in this experiment.

Model	Subset S_1			Subset S_2			Subset S_3		
	Total	FP	FN	Total	FP	FN	Total	FP	FN
Window model	41	5	2	35	3	1	44	5	1
K-means	66	35	7	64	40	9	73	42	9
Box model	22	0	16	15	0	18	25	0	15

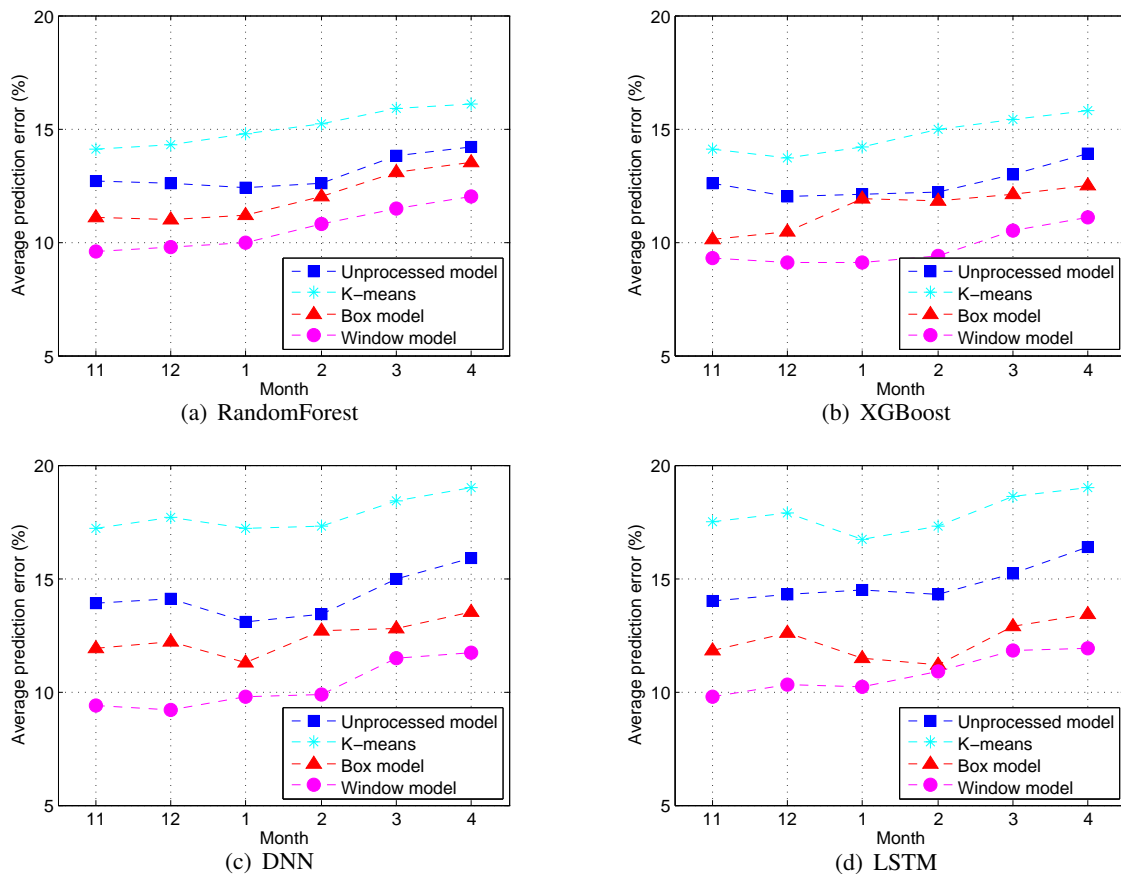


Figure 4. Average prediction error rate of different data preprocessing models in **heating season**. In this experiment, four existing load data prediction schemes, RandomForest, XGBoost, DNN and LSTM, are used to provide the experimental results.

A series of experiments are performed to test the accuracy that three preprocessing models detect abnormal values over the same sample subset. Two different measure standards, false positive (FP) and false negative (FN), are used to evaluate the performance of different preprocessing models, where

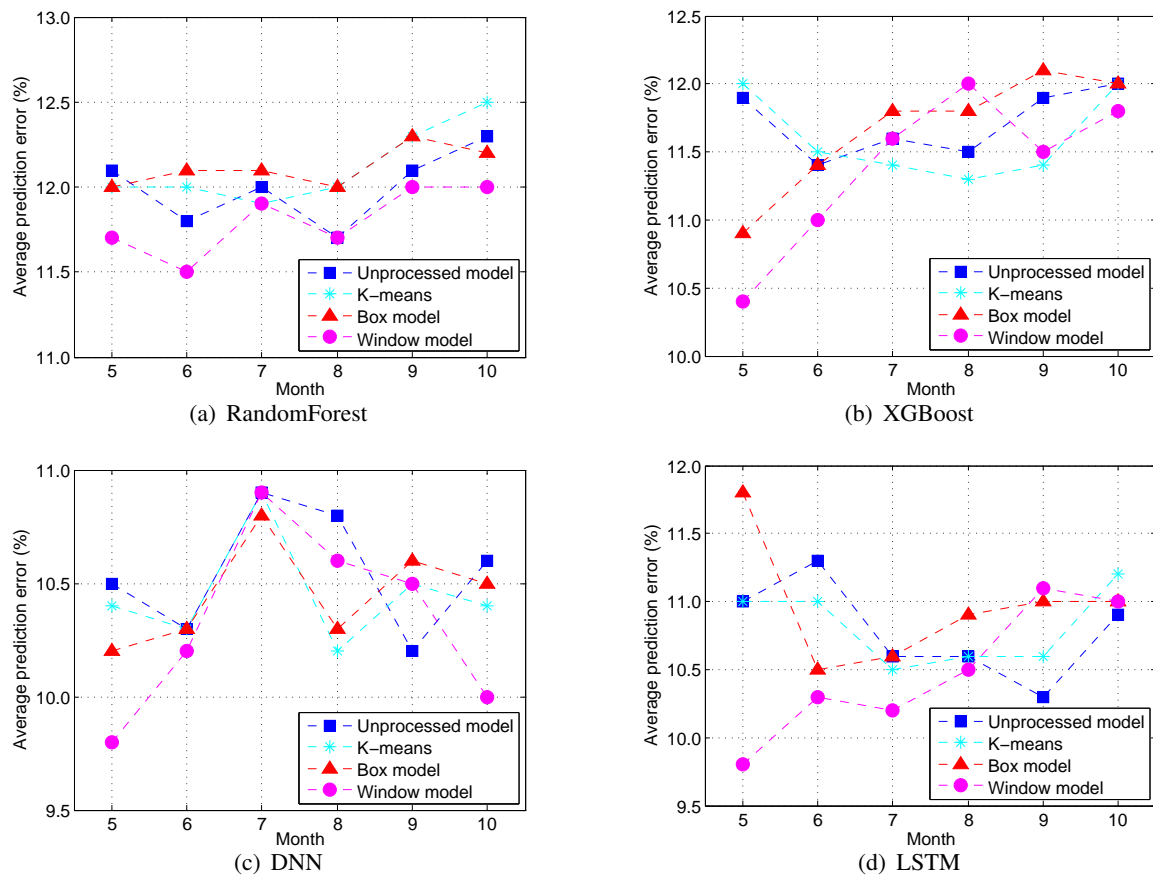


Figure 5. Average prediction error rate of different data preprocessing models in **non-heating season**. In this experiment, four existing load data prediction schemes, RandomForest, XGBoost, DNN and LSTM, are used to provide the experimental results.

false positive indicates the number that the normal sample is detected as abnormal sample, while false negative indicates the number that the abnormal sample is detected as normal sample.

The experimental results are shown in Table 2. It can be seen that the window model always provides a lowest error number (sum of FP and FN) in three preprocessing models. This demonstrates that proposed window model has a higher detection capability for abnormal samples. In contrast, K-means model always detects the most abnormal samples, whatever sample subset is used, and it also give a higher false positive number, which implies that K-means model may be more lenient for abnormal data. This is mainly because K-means model always divides the normally increasing data (including a lot of normal values) into two categories during the clustering process, resulting in some normal data are consistently misjudged as abnormal values. In addition, the Box model is rather conservative due to a zero false positive number, because the Box model detect the abnormal value by the distance between the upper and lower limits of the given gas volume and the median. Therefore, there is usually no misjudgment for the normal value.

In order to further show the advantages of proposed window model, we test the overall performance of four data preprocessing models, window model, K-means model, Box model, and

Unprocessed model over four load data prediction schemes, RandomForest [25], eXtreme Gradient Boosting (XGBoost) [26], Deep Neural Network (DNN) [8], Long Short Term Memory (LSTM) [27], where Unprocessed model means that original data will not be preprocessed. In the experiments, the total load data set from 2017 to 2019 is used and the detected abnormal data is uniformly completed by the adjacent mean interpolation method. The fluctuation deviation is fixed as $E = 0.3^{\ddagger}$ in proposed window model.

Figures 4 and 5 show the average prediction error rate of different data pre-processing models in heating season and non-heating season, respectively. In these figures, the horizontal axis represents the months of heating season and non-heating season, and the vertical axis represents the average prediction error rate. It is easy to observe that our proposed window model can obtain a lowest average prediction error rate in heating season, whatever data prediction models are used. Specifically, the average reduction of window model is more than 4–5% for K-means model, 1–2% for Box model, and 3% for Unprocessed model. These results indicate that the proposed window model can maintain a high detection capability. Furthermore, we can observe that the average prediction error rate of K-means model is significantly higher than that of the Unprocessed model, e.g., approximately 1.5% for RandomForest, 2.2% for XGBoost, 3.0% for DNN and 2.8% for LSTM. This implies that when the original data is processed by K-means model, the average error rate increases instead. This interesting phenomenon can be explained easily. Because K-means model decides too many normal values as abnormal data, these misjudged data just seriously affect the training process of different prediction algorithms, resulting in a large deviation between the predicted value and the original value.

In addition, we also see that the advantage of proposed model in non-heating season is not obvious. This is because the load data in non-heating season has only a slight fluctuation, resulting in very few abnormal data. Accordingly, the advantage of eliminating abnormal data for window model is difficult to play.

5.2. Test for ensemble multilayer perceptron model

In order to show the advantages of our proposed ensemble MLP method, we firstly test the impact of the number of multilayer perceptron in ensemble model. In this test, we use three data preprocessing models, Window model, Box model, and Unprocessed model, to clean the original data, and then implement the ensemble model with different number of single multilayer perceptron.

Figure 6 shows the average prediction error rate when the number of multilayer perceptron c is set to 1, 2, 3, \dots , 12, respectively. As can be seen from the figure, the average prediction error rate is tending towards stability with an increasing c , and does not change seriously when $c > 10$. In general, the larger the number of multilayer perceptron, the lower the average prediction error rate, whatever data preprocessing model is used. This means that the ensemble model will not always gain benefits with the number of MLP increasing. In fact, we can explain this phenomenon that since single MLP can obtain different prediction results from each other so that the ensemble effect is similar to majority voting [24]. The prediction capability is significantly improved accordingly. Nevertheless, with the number of MLP increasing, the diversity of single MLP will reduce and even disappear so that ensemble model tends to generate the same results.

In addition, we also test the performance of adaptive weight correction. In order to give a more insight, we compare the performance of ensemble model with and without adaptive weight correction.

[‡]This is an experience value for many natural gas enterprises.

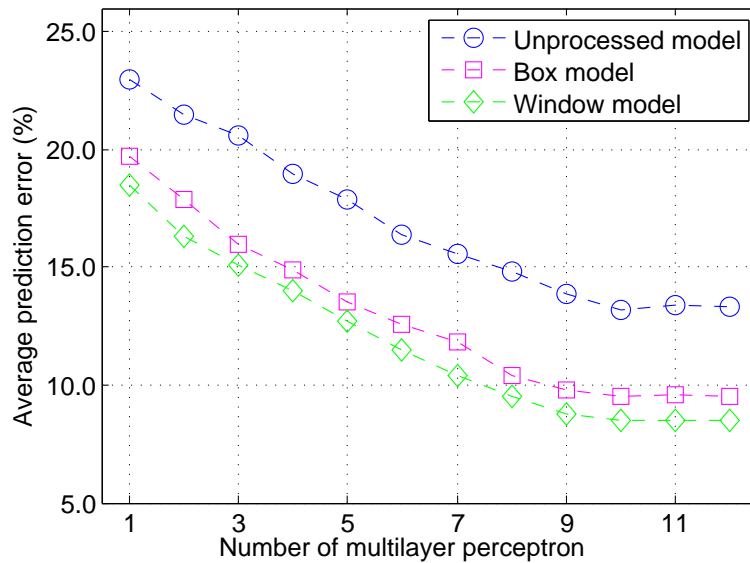


Figure 6. Average prediction error rate for different number c of multilayer perceptron in proposed ensemble model. In this experiment, three data preprocessing model, Unprocessed model, Box model and proposed Window model, are used to give the experimental results.

In the experiments, the total data set is used to test experimental results. All data are pre-processed with three preprocessing models, Window model, Box model, and Unprocessed model. The ensemble model is employed to give a fair comparison.

We test the average prediction error rate of ensemble MLP model with and without adaptive weight correction. Table 3 and Table 4 give the experimental results in heating season and non-heating season, respectively. In our experiments, the number of MLP is set to $c = 1, 5, 10$. It is easy to observe that for the heating season, when adaptive weight correction is adopted in proposed ensemble MLP model, the average prediction error rate are slightly lower than the case without adaptive weight correction. The average reductions of PSNR are more than 2.1–6.0% for $c = 1$, 2.7–7.8% for $c = 5$, and 1.6–6.6% for $c = 10$, respectively. That is mainly because proposed adaptive weight correction dynamically modifies the weight ratio of the prediction results according to the changes in gas load data, resulting in a higher prediction accuracy. Moreover, for non-heating season, the average reductions of PSNR is not much, and most of them are maintained at 1–1.5%.

5.3. Performance comparison with state-of-the-arts

In this section, we compare the proposed scheme with existing four load data forecasting schemes, RandomForest-based scheme [25], XGBoost-based scheme [26], DNN-based scheme [8], LSTM-based scheme [27]. In this experiment, for LSTM, we set the number of hidden layer as 3 and the nodes of each layer as 10, for RandomForest and XGBoost, the number of sub-model is fixed as the range [50, 200], for DNN, the number of hidden layer is set to 20. For proposed scheme, the window model is used and the ensemble number of MLP is $c = 10$. We perform crossing segmentation validation over the total data set and divide it as two parts: one is used as training set and another is testing set. All experiments are implemented ten times to give an average result.

Table 3. *APEE* performance comparison of proposed ensemble MLP model with and without adaptive weight correction over the total database. In this test, the experiment is implemented over the data of **heating season** and the number of MLP are set to $c = 1$, $c = 5$ and $c = 10$, respectively.

Month	Ensemble without correction			Ensemble with correction		
	$c = 1$	$c = 5$	$c = 10$	$c = 1$	$c = 5$	$c = 10$
11	21.5%	16.7%	14.5%	18.7%	13.3%	10.1%
12	23.5%	16.9%	13.6%	21.4%	14.1%	9.3%
1	29.3%	21.2%	19.1%	23.3%	16.5%	12.3%
2	31.2%	25.4%	19.8%	25.3%	17.6%	13.2%
3	20.6%	15.9%	13.2%	17.6%	11.1%	9.7%
4	19.5%	13.7%	11.4%	16.9%	11.0%	9.8%

Table 4. *APEE* performance comparison of proposed ensemble MLP model with and without adaptive weight correction over the total database. In this test, the experiment is implemented over the data of **non-heating season** and the number of MLP are set to $c = 1$, $c = 5$ and $c = 10$, respectively.

Month	Ensemble without correction			Ensemble with correction		
	$c = 1$	$c = 5$	$c = 10$	$c = 1$	$c = 5$	$c = 10$
5	20.5%	17.0%	13.8%	18.9%	16.5%	13.0%
6	19.5%	16.2%	14.2%	19.0%	15.6%	14.0%
7	19.5%	15.0%	13.5%	18.3%	14.5%	13.5%
8	19.0%	15.1%	13.6%	18.9%	15.1%	13.5%
9	20.1%	15.0%	13.5%	18.9%	14.9%	12.9%
10	19.6%	15.5%	13.0%	18.5%	14.8%	12.6%

Firstly, we test the *MAE* and *RMSE* values by comparing proposed scheme with existing four load data forecasting schemes. The results are shown in Table 5. Actually, we can observe from experimental results that, the prediction performance of proposed EMLP scheme has an obvious improvement in heating season, whichever evaluation metric is used (Corresponds to smaller values comparing with other four schemes in Table 5). This demonstrates that proposed scheme has a superior performance comparing with other existing schemes. Moreover, we can see that for *MAE* metric, the prediction performance of proposed EMLP scheme has relatively stable prediction results in heating season, while for *RMSE* metric, the proposed EMLP scheme shows a slight prediction fluctuation. This mainly because for *RMSE* metric, the error calculations are all squared values, slight data fluctuations may lead to larger prediction errors. In addition, we can also observe that proposed scheme always has stable higher prediction accuracy, no matter in heating season or

non-heating season. This implies that proposed scheme has a stronger generalization performance.

Table 5. MAE and RMSE performance comparison for five forecasting schemes. The experiment tests the heating season and non-heating season, respectively. In this test, the experiment is implemented over the data of **heating season** and **non-heating season**, respectively.

Metrics	Existing schemes	Heating season						Non-heating season					
		11	12	1	2	3	4	5	6	7	8	9	10
MAE	EMLP	7.8	7.3	7.0	7.9	8.1	8.2	6.0	6.2	6.6	6.4	6.8	6.6
	LSTM	8.4	8.2	7.1	8.5	9.4	10.2	6.8	6.6	6.9	6.6	6.1	6.9
	DNN	8.9	8.3	7.9	8.4	9.9	10.6	6.2	6.1	6.3	6.1	6.3	6.1
	RandomForest	7.7	8.1	7.1	8.4	9.1	8.6	6.2	6.6	6.4	6.5	6.0	6.9
	XGBoost	7.3	8.0	7.2	8.1	8.8	8.7	6.0	7.0	6.5	6.6	6.4	6.8
RMSE	EMLP	10.3	11.7	10.6	11.9	13.1	12.9	9.0	8.2	7.6	7.7	7.8	8.6
	LSTM	11.4	13.2	12.1	13.8	17.1	16.2	8.8	7.6	7.4	8.1	7.9	9.0
	DNN	12.9	13.3	11.9	14.0	17.5	16.9	8.4	7.1	7.2	7.8	8.3	8.8
	RandomForest	11.0	12.0	11.2	12.6	14.2	14.2	9.2	9.6	8.1	8.9	8.3	8.9
	XGBoost	10.7	12.1	11.2	12.1	14.1	14.7	9.2	9.3	8.5	8.5	8.4	8.5

In order to give more insight, Figure 7 shows the average prediction error rate for five different prediction schemes in heating season and non-heating season, respectively. As can be seen from this figure, the proposed scheme has an obvious advantage in heating season for gas daily load forecasting. We can explain easily this phenomenon as follows. Proposed method firstly eliminates abnormal data with large fluctuations through window model. Subsequently, ensemble model integrates multiple MLP sub-neural networks to further suppresses the over-fitting phenomenon, and finally significantly improves the accuracy of data prediction. Moreover, proposed adaptive weight correction method dynamically adjusts the weight value of the subsequent data by calculating the deviation of the forward data. This processing leads to a positive feedback so that the overall performance of proposed scheme has an obvious improvement comparing with the other existing four schemes. In addition, we should also note that the prediction performance of proposed scheme has only a slight advantage in non-heating season. This is mainly because that the gas consumption is obviously lower in the non-heating season so that the fluctuation of the daily load data is smaller, and thus the performance advantages of proposed method cannot be fully reflected. On the other hand, the small fluctuation also makes the training set and testing set are more similar, which lets these non-ensemble methods, e.g., DNN-based scheme and LSTM-based scheme, are more prone to overfitting.

Furthermore, in order to give more insight, we employ genetic algorithm to implement the hyperparameter optimization for each scheme and form the optimal prediction model. In order to ensure a fair comparison, each experiment is repeated ten times to provide the average values. The experimental results are shown in Figure 8. We can find that after hyperparameter optimization, the prediction errors for proposed scheme, RandomForest-based scheme and XGBoost-based scheme have a slight decreasing. This means that hyperparameter optimization can bring some advantage for

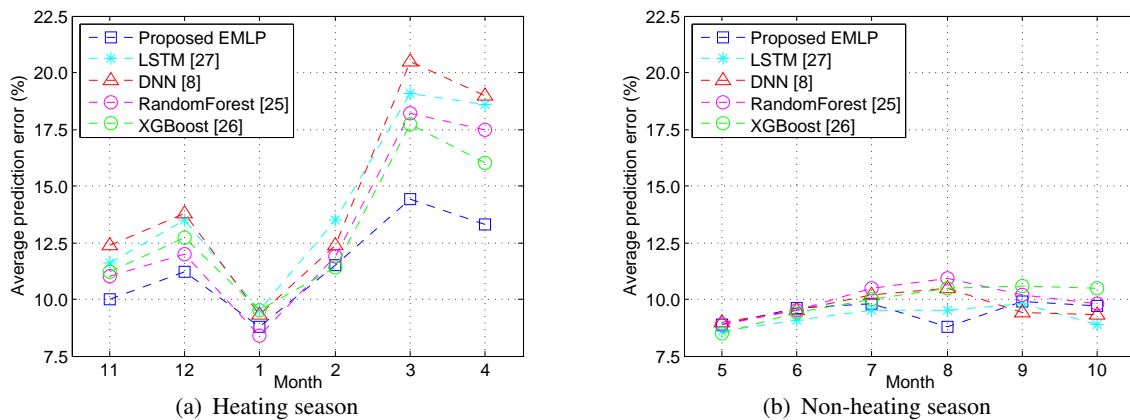


Figure 7. Average prediction error rate without hyperparameter optimization for five state-of-the-art prediction schemes in (a) heating season and (b) non-heating season.

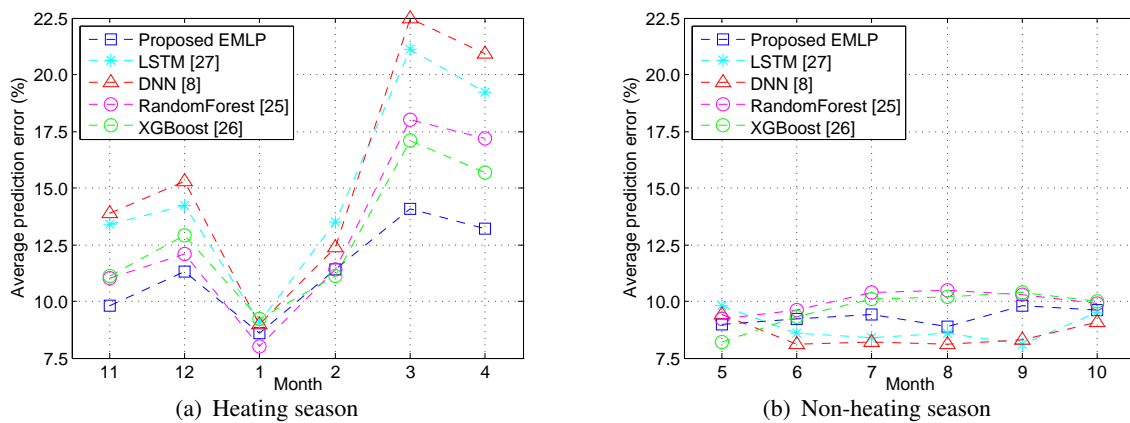


Figure 8. Average prediction error rate with hyperparameter optimization for five state-of-the-art prediction schemes in (a) heating season and (b) non-heating season.

these three schemes. Nevertheless, parameter optimization are not very obvious for DNN-based scheme and LSTM-based scheme, even the error rate has increased. This is mainly because for DNN-based scheme and LSTM-based scheme, hyperparameter optimization strengthens the learning aiming at historical data rules, which can lead to greater deviations when some new data are predicted. In contrast, RandomForest-based scheme and XGBoost-based scheme belong to ensemble learning schemes. Compared with neural networks, ensemble learning schemes can alleviate the problem of overfitting. For proposed scheme, on the basis of ensemble learning, we use dynamic weight to further reduce the impact of overfitting and adapt to the changes in new data, resulting in an overall performance improvement in prediction accuracy.

6. Conclusions and future works

Load forecasting for natural gas is a much needed feature in the presence of smart city construction. This important requirement, however, is largely ignored in existing gas forecasting

model, because they mostly relies on a combined forecasting model by simply integrating multiple single-forecasting models, and may obtain accordingly an inferior prediction performance due to redundant single-forecasting model. We filled the gap by designing a new gas load forecasting scheme based on ensemble multilayer perceptron (EMLP) with adaptive weight correction. This new scheme has a significant advantage that it can integrate multiple weak multilayer perceptron to give a more accurate prediction result.

Our method firstly normalizes multi-source data as original data set, and then segment it by designing a window model to extract the abnormal values, which are interpolated to form a complete normalized data set. Subsequently, a series of multilayer perceptron (MLP) networks are integrated to construct an ensemble forecasting model. A weight correction function is further introduced to dynamically modify the weight of the prediction result. Extensive experiments demonstrate that compared with existing short-term forecasting methods, our method can accurately forecast the daily gas load and outperforms the state-of-the-art in terms of the prediction accuracy.

Finally, it is still an open research challenge to design gas load forecasting model. As a future work, we will investigate forecasting model to further improve prediction accuracy while maintaining better forecasting stability.

Acknowledgments

This work was supported by the Natural Science Foundation of Shanghai (20ZR1421600).

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this article.

References

1. H. Jiang, K. Wang, Y. Wang, M. Gao, Y. Zhang, Energy big data: A survey, *IEEE Access*, **4** (2016), 3844–3861.
2. Z. Ma, J. Xie, H. Li, Q. Sun, Z. Si, J. Zhang, et al., The role of data analysis in the development of intelligent energy networks, *IEEE Network*, **31** (2017), 88–95.
3. M. Fagiani, S. Squartini, L. Gabrielli, S. Spinsante, F. Piazza, A review of datasets and load forecasting techniques for smart natural gas and water grids: Analysis and experiments, *Neurocomputing*, **170** (2015), 448–465.
4. O. Laib, M. T. Khadir, L. Mihaylova, *A Gaussian process regression for natural gas consumption prediction based on time series data*, 2018 21st International Conference on Information Fusion (FUSION), IEEE, 2018.
5. J. Ravnik, M. Hribersek, A method for natural gas forecasting and preliminary allocation based on unique standard natural gas consumption profiles, *Energy*, **180** (2019), 149–162.
6. W. Qiao, Z. Yang, Z. Kang, Z. Pan, Short-term natural gas consumption prediction based on volterra adaptive filter and improved whale optimization algorithm, *Eng. Appl. Artif. Intell.*, **87** (2020), 103323.

7. N. Wei, C. Li, X. Peng, Y. Li, F. Zeng, Daily natural gas consumption forecasting via the application of a novel hybrid model, *Appl. Energy*, **250** (2019), 358–368.
8. G. D. Merkel, R. J. Povinelli, R. H. Brown, Short-term load forecasting of natural gas with deep neural network regression, *Energies*, **11** (2018), 2008.
9. B. Ervural, O. Beyca, S. Zaim, Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption, *Procedia Soc. Behav. Sci.*, **235** (2016), 537–545.
10. I. Panapakidis, A. Dagoumas, Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model, *Energy*, **118** (2017), 231–245.
11. W. Qiao, K. Huang, M. Azimi, S. Han, A novel hybrid prediction model for hourly gas consumption in supply side based on improved whale optimization algorithm and relevance vector machine, *IEEE Access*, **7** (2019), 88218–88230.
12. J. Lei, T. Jin, J. Hao, F. Li, Short-term load forecasting with clustering-regression model in distributed cluster, *Cluster Comput.*, **22** (2019), 10163–10173.
13. L. Wang, S. Mao, B. M. Wilamowski, R. M. Nelms, Ensemble learning for load forecasting, *IEEE Trans. Green Commun. Networking*, **4** (2020), 616–628.
14. T. Li, Y. Wang, N. Zhang, Combining probability density forecasts for power electrical loads, *IEEE Trans. Smart Grid*, **11** (2020), 1679–1690.
15. C. Feng, M. Sun, J. Zhang, Reinforced deterministic and probabilistic load forecasting via Q -learning dynamic model selection, *IEEE Trans. Smart Grid*, **11** (2020), 1377–1386.
16. W. Zheng, Z. Li, X. Liang, J. Zheng, Q. H. Wu, F. Hu, Decentralized state estimation of combined heat and power system considering communication packet loss, *J. Mod. Power Syst. Clean Energy*, **8** (2020), 646–656.
17. B. Zhang, C. Dou, D. Yue, Z. Zhang, T. Zhang, A packet loss-dependent event-triggered cyber-physical cooperative control strategy for islanded microgrid, *IEEE Trans. Cybern.*, **51** (2021), 267–282.
18. F. Yu, X. Xu, A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network, *Appl. Energy*, **134** (2014), 102–113.
19. Y. Karadede, G. Ozdemir, E. Aydemir, Breeder hybrid algorithm approach for natural gas demand forecasting model, *Energy*, **141** (2017), 1269–1284.
20. F. Taspinar, N. Celebi, N. Tutkun, Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods, *Energy Build.*, **56** (2013), 23–31.
21. G. Zu, L. Lu, X. Xu, Study of information entropy combination model for short-term gas load prediction, *Comput. Appl. Software*, **30** (2013), 129–131.
22. J. Tang, C. Deng, G. B. Huang, Extreme learning machine for multilayer perceptron, *IEEE Trans. Neural Networks Learn. Syst.*, **27** (2016), 809–821.
23. M. Walker, Y. Dovoedo, S. Chakraborti, C. W. Hilton, An improved boxplot for univariate data, *Am. Stat.*, **72** (2018), 348–353.

24. F. Li, K. Wu, J. Lei, M. Wen, Z. Bi, C. Gu, Steganalysis over large-scale social networks with high-order joint features and clustering ensembles, *IEEE Trans. Inf. Forensics Secur.*, **11** (2016), 344–357.
25. C. Li, Y. Tao, W. Ao, S. Yang, Y. Bai, Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition, *Energy*, **165** (2018), 1220–1227.
26. T. Chen, C. Guestrin, *Xgboost: A scalable tree boosting system*, Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016.
27. A. Sagheer, M. Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks, *Neurocomputing*, **323** (2019), 203–213.
28. S. Chen, Y. Wu, B. Luk, Combined genetic algorithm optimization and regularized orthogonal least squares learning for radial basis function networks, *IEEE Trans. Neural Networks*, **10** (1999), 1239–1243.
29. X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, et al., Multiple kernel k -means with incomplete kernels, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 1191–1204.
30. F. Li, G. Zhou, J. Lei, Reliable data transmission in wireless sensor networks with data decomposition and ensemble recovery, *Math. Biosci. Eng.*, **16** (2019), 4526–4545.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)