2021 IEEE 7th International Conference on Control Science and Systems Engineering (ICCSSE) | 978-1-6654-4405-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCSSE52761.2021.9545089

Ensemble Online Sequential Extreme Learning Machine for Air Quality Prediction

Ye Liu
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, China
ly@szu.edu.cn

Weipeng Cao*
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, China
caoweipeng@szu.edu.cn

Yiwen Liu
College of International Exchange
Shenzhen University
Shenzhen, China
lyw@szu.edu.cn

Dachuan Li
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
lidc3@mail.sustech.edu.cn

Qiang Wang
Department of Computer Science and Engineering
Southern University of Science and Technology
Shenzhen, China
wangq8@sustech.edu.cn

***Abstract*—Online Sequential Extreme Learning Machine (OS-ELM) has been confirmed by numerous studies to be an effective algorithm for online learning scenarios. However, we found that some parameters of OS-ELM are randomly assigned and remain unchanged in the subsequent learning process, which leads to great instability in the model performance in practice. To alleviate this problem, we propose a novel ensemble OS-ELM algorithm (EOS-ELM-R) for solving air quality prediction problems. EOS-ELM-R uses multiple distribution functions to initialize the random parameters of the base OS-ELM models and its final output is the average of the predictions of these base models. Extensive experimental results on two real-world air quality prediction problems show that EOS-ELM-R is effective, and it can achieve better generalization capabilities than similar algorithms.**

Keywords*—Online learning, online sequential extreme learning machine, extreme learning machine, ensemble learning, air quality prediction

I. Introduction

Air quality has a direct and significant impact on human health. With the advancement of science and technology, more and more countries have begun to invest huge resources to build air quality monitoring systems, and formulate relevant policies based on these data feedbacks to protect the health of their citizens. In recent years, as machine learning technology (especially deep learning) has shown exciting breakthroughs in many fields such as computer vision and natural language processing [1][2], researchers have begun to try to build air quality prediction systems based on related technologies. Some representative work is as follows:

Kök I et al. [3] proposed to use Long Short Term Memory (LSTM) networks to predict the air quality in a smart city. They used several experimental results to demonstrate the feasibility of this method. In [4], Wang J et al. designed a deep Spatial-Temporal Ensemble (STE) model to evaluate the air quality of Beijing, China. STE is essentially a combination of ensemble learning, causal analysis, and deep LSTM technology. Considering the spatial correlations among air pollutants, Yi X et al. [5] used a spatial transformation module to transform the spatial sparse data into the consistent input and then used a deep distributed network to fuse the heterogeneous urban data. In [6], Gu K et al. proposed an improved Support Vector Machine (SVM) to construct the air quality evaluation model. They also used partial least squares to analyze the importance of different pollution factors to air quality. For more cases of using machine learning technology to model air quality prediction problems, please refer to [7].

The above work has demonstrated the great potential of using machine learning technology to build the air quality prediction system. However, the above models face a common shortcoming: the training efficiency of the models is low. The reason for this phenomenon is due to the training mechanism of traditional deep learning techniques such as LSTM and traditional statistical machine learning methods such as SVM. For example, traditional neural networks such as LSTM use the error back propagation method to iteratively finetune all parameters of the model, which determines that the training process of the model is time-consuming.

To improve the training efficiency of the model, in recent years, researchers try to use some neural networks that adopt non-iterative training mechanisms such as Extreme Learning Machine (ELM) [8] to model air quality prediction problems. Taking an ELM with a single hidden layer as an example, its network structure is shown in Fig. 1. In its training process, the input weights between the input layer and the hidden layer and the thresholds of the hidden nodes are randomly assigned, and these parameters remain unchanged in the subsequent model training process. ELM only needs to use the least square method

This work was supported by the National Natural Science Foundation of China (Grant No. 61836005) and Shenzhen Basic Research Fund (Grant No. JCYJ20200813091134001). *Corresponding author.

to calculate its output weights between the hidden layer and the output layer to complete the model training. This non-iterative training mechanism makes the training efficiency of ELM significantly higher than the above traditional methods. Some representative work based on ELM to build the air quality prediction system includes: Jiang F et al. [9] proposed a hybrid learning method to predict the urban air quality index (AQI). In their method, they used Wavelet Packet Decomposition (WPD) to preprocess the original AQI data and then used an improved ELM algorithm to train the prediction model. Considering the temporal and spatial characteristics of air quality data, Liu B et al. [10] proposed a Spatio-Temporal ELM (STELM) method to utilize this feature to make a more accurate prediction.

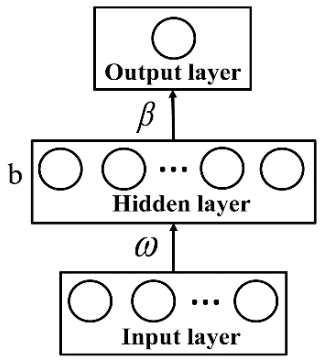

Fig. 1. The network structure of ELM.

Although the above ELM-based air quality prediction models show great potential, these algorithms use the batch learning training mode. This means that when new training data arrives, one needs to merge them into the existing training data before using them to re-train the model. With the accumulation of data, the training of these models will consume a lot of computing resources.

To avoid the above problem, we choose the OS-ELM algorithm [11], a variant of ELM, to model the air quality prediction problems, which is an effective incremental learning algorithm for online learning scenarios. The network structure of OS-ELM is the same as ELM, as shown in Fig. 1. The training process of OS-ELM includes two stages: initialization and sequential learning. The training mechanism of its initialization stage is the same as that of ELM, while its sequential learning stage does not use the learned data and only uses new data to update the model parameters. OS-ELM can avoid the problem of insufficient computing resources caused by data accumulation.

However, in the initialization stage of OS-ELM, its input weights and hidden biases are randomly assigned based on a uniform distribution, and these parameters remain unchanged in the subsequent updating process. Once the initialization quality of these parameters is not good, the performance of the model will face great uncertainty. To solve this problem, a novel ensemble OS-ELM algorithm (i.e., EOS-ELM-R) is proposed to model regression problems. In EOS-ELM-R, the base learners are OS-ELMs. We use a variety of distribution functions to initialize these base learners separately, thereby increasing the diversity of the base models. The final output of EOS-ELM-R is the average of the predicted values of these base models.

The main contributions of this paper can be summarized as follows:

(1) This paper proposes a novel ensemble OS-ELM algorithm for modeling regression problems;

(2) The proposed EOS-ELM-R can effectively alleviate the problem of large prediction deviations caused by the poor quality of random parameters in OS-ELM;

(3) We apply EOS-ELM-R to model the air quality prediction problem, and the experimental results show the effectiveness of the proposed method.

The remainder of this paper is organized as follows: In Sec. II, we introduce the training mechanism of OS-ELM, so that readers can better understand our proposed EOS-ELM-R algorithm. In Sec. III, we give the details of EOS-ELM-R including its learning pipeline and pseudo-code. In Sec. IV, we give the experimental results of EOS-ELM-R on air quality prediction problems. We summarize the contribution of this paper in Sec. V.

II. Related Work

In this section, we introduce the training mechanism of OS-ELM which is directly related to the proposed algorithm.

OS-ELM is a single hidden layer neural network for online sequential learning scenarios, and its network structure is the same as ELM, as shown in Fig. 1. In Fig. 1, ω and b, i.e., the input weights and hidden biases, are generated randomly from [-1, 1] under a uniform distribution, which will be kept frozen throughout the remaining training process of the model. The pseudo-code of OS-ELM is described in algorithm 1.

Algorithm 1: OS-ELLM

Initialization phase:

Given an initial training dataset $D_0=\{(x_j,t_j \mid x_j\in R^d, t_j=R^m)\}_{j=1}^{N_0}$, then use the following method to train an initial model:

(1) Set $H_0\beta_0=T_0$, then $\beta_0=H_0^{+}T_0$

where $$H_0=\begin{pmatrix} G(\omega_1\cdot x_1+b_1) & \dots & G(\omega_L\cdot x_1+b_L) \\ \vdots & \ddots & \vdots \\ G(\omega_1\cdot x_{N_0}+b_1) & \cdots & G(\omega_L\cdot x_{N_0}+b_L) \end{pmatrix}_{N_0\times L},$$

$$\beta_0=\begin{bmatrix}\beta_1^T\\ \vdots \\ \beta_L^T\end{bmatrix}_{L\times m}, T_0=\begin{bmatrix} t_1^T \\ \vdots \\ t_{N_0}^T\end{bmatrix}_{N_0\times m},$$ L refers to the number of hidden nodes, and H_0^{+} refers to the Moore–Penrose generalized inverse of H_0.

(2) Let $k=0$ and $P_0=(H_0^TH_0)^{-1}$.

Online sequential learning phase:

(3) Given the $(k+1)_{th}$ chunk of new dataset D_{k+1}, update the

output of the hidden layer: $H_{k+1} = [H_k^T, H_{k+1}^T]^T$.

(4) Update the output weights: $\beta_{k+1} = \beta_k + P_{k+1}H_{k+1}^T(T_{k+1} - H_{k+1}\beta_k)$. Here $P_{k+1} = P_k - P_k H_{k+1}^T (I + H_{k+1} P H_{k+1}^T)^{-1} H_{k+1} P_k$ and T_{k+1} denote the labels of the new samples.

(5) Set $k = k+1$. Go back to step (3) until all the new data are learned.

III. The Details of EOS-ELM-R

From algorithm 1, one can infer that the quality of the random parameters (i.e., input weights and the hidden biases) has a direct impact on the performance of OS-ELM. Unfortunately, there is no universal and effective method to guarantee the quality of these random parameters, which leads to the instability of OS-ELM in practice.

Inspired by ensemble learning, we propose a novel ensemble OS-ELM (i.e., EOS-ELM-R) to alleviate the above problem. The learning pipeline of EOS-ELM-R is shown in Fig. 2. From Fig. 2, one can observe that EOS-ELM-R contains multiple base OS-ELMs, and each base model is initialized with a specific distribution. The final output of EOS-ELM-R is the average of the outputs of base models. In this way, EOS-ELM-R avoids the problem of large prediction deviation caused by the poor quality of random initialization of a single model.

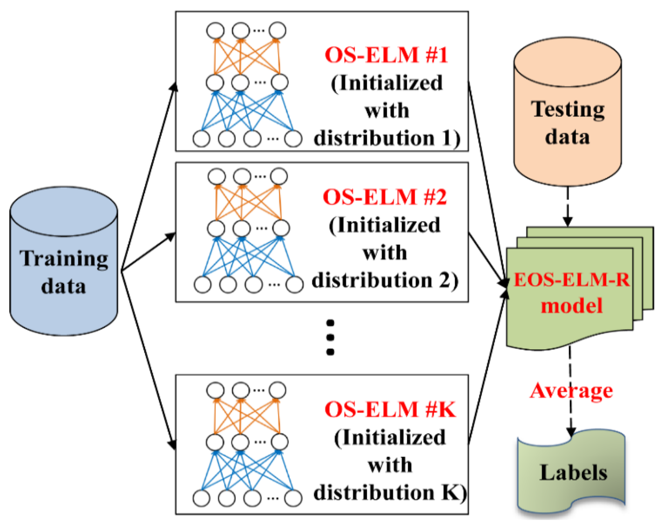

Fig. 2. The learning pipeline of EOS-ELM-R.

The pseudo-code of EOS-ELM-R is shown in algorithm 2.

Algorithm 2 EOS-ELM-R

Initialization phase:

Given an initial training dataset $D_0 = \{(x_j, t_j \mid x_j \in R^d, t_j = R^m)\}_{j=1}^{N_0}$ and **K** distribution functions, such as Uniform, Gaussian, and Gamma distributions. Initialize each base OS-ELM with each distribution function and obtain the corresponding base models according to the training mechanism mentioned in algorithm 1.

Online sequential learning phase:

When new data arrives, update all the initial models until all the new data are learned. This updating process is the same as the online sequential learning phase of algorithm 1.

Decision-making phase:

When a testing sample arrives, use the base models to predict its label and then obtain the final label by averaging the predictions of the base models.

IV. Experiments and Results Analysis

In this section, we evaluate the performance of EOS-ELM-R through two air quality prediction problems. Our experiments are conducted with MATLAB R2018b on Mac OS with Intel Core i7 CPU and 16 GB RAM.

A. Dataset

The concentrations of Particulate Matter 2.5 (PM 2.5) and PM10 in the air are currently the most important indicators used by many governments to evaluate air quality.

In this paper, we choose the air quality data published by the governments of Beijing and Oslo to predict their air pollution based on the proposed model. For the data from Beijing, we predict the concentration of PM 2.5 in the air, while for the data from Oslo, we predict the concentration of PM 10 in the air. These two public datasets can be obtained from [12] and the website http://lib.stat.cmu.edu/datasets/, respectively. Additional information about these two datasets can be found in Table I. We use the same method as [13] to preprocess these two datasets.

TABLE I. Details of Experimental Datasets

Dataset	Attributes	Samples
Beijing PM 2.5	11	41757
Oslo PM 10	7	500

B. Experimental Setting

In our experiments, we choose *Sigmoid* function (i.e., $G(z) = 1/(\exp(-z))$) as the activation function of EOS-ELM-R and the comparison algorithms OS-ELM and EOS-ELM [14]. The number of hidden nodes in these algorithms is set to 30. We set the number of distribution functions in EOS-ELM-R to 3, and choose the Uniform, Gaussian, and Gamma distributions to initialize the base models of EOS-ELM-R.

For each problem, the original dataset is split into a training set and a testing set according to the division ratio 7:3. The training set is further split into an initial set and a sequential set according to the division ratio 3:7 for simulating the learning process of online learning. Then we simulate the three most common scenarios: learning new samples one-by-one, learning new samples chunk-by-chunk (chunk size is fixed and bigger than one), and learning new samples chunk-by-chunk (chunk size is random).

We chose Root Mean Square Error (RMSE) as the indicator to evaluate the model performance, which can be obtained by using the following equation:

$$RMSE=\sqrt{\sum_{i=1}^{N}\frac{(t_i-y_i)^2}{N}} \tag{1}$$

where t_i is the model's prediction label for the i-th sample, y_i is the real label of the i-th sample, and N is the number of samples.

C. Experimental Results

The experimental results are shown in Tables II-IV, which are the average of the results over 20 trials. Note that the smaller the RMSE, the better the model performance.

TABLE II. THE EXPERIMENTAL RESULTS OF OS-ELM, EOS-ELM, AND EOS-ELM-R (ONE-BY-ONE)

Datasets	**RMSE**	**OS-ELM**	**EOS-ELM**	**EOS-ELM-R**
Beijing PM 2.5	Training	0.0773	0.0767	0.0763
	Testing	0.0777	0.0771	**0.0756**
Oslo PM 10	Training	0.1475	0.1456	0.1442
	Testing	0.1646	0.1617	**0.1602**

TABLE III. THE EXPERIMENTAL RESULTS OF OS-ELM, EOS-ELM, AND EOS-ELM-R (CHUNK-BY-CHUNK WITH A FIXED SIZE: 20)

Datasets	**RMSE**	**OS-ELM**	**EOS-ELM**	**EOS-ELM-R**
Beijing PM 2.5	Training	0.0754	0.0746	0.0736
	Testing	0.0787	0.0779	**0.0767**
Oslo PM 10	Training	0.1575	0.1546	0.1563
	Testing	0.1654	0.1644	**0.1609**

TABLE IV. THE EXPERIMENTAL RESULTS OF OS-ELM, EOS-ELM, AND EOS-ELM-R (CHUNK-BY-CHUNK WITH A RANDOM SIZE: [10, 20])

Datasets	**RMSE**	**OS-ELM**	**EOS-ELM**	**EOS-ELM-R**
Beijing PM 2.5	Training	0.0773	0.0759	0.0748
	Testing	0.0777	0.0765	**0.0754**
Oslo PM 10	Training	0.1675	0.1675	0.1533
	Testing	0.2002	0.1908	**0.1686**

To show the experimental results more intuitively, we visualize the results in Tables II-IV as Fig. 3 and Fig. 4.

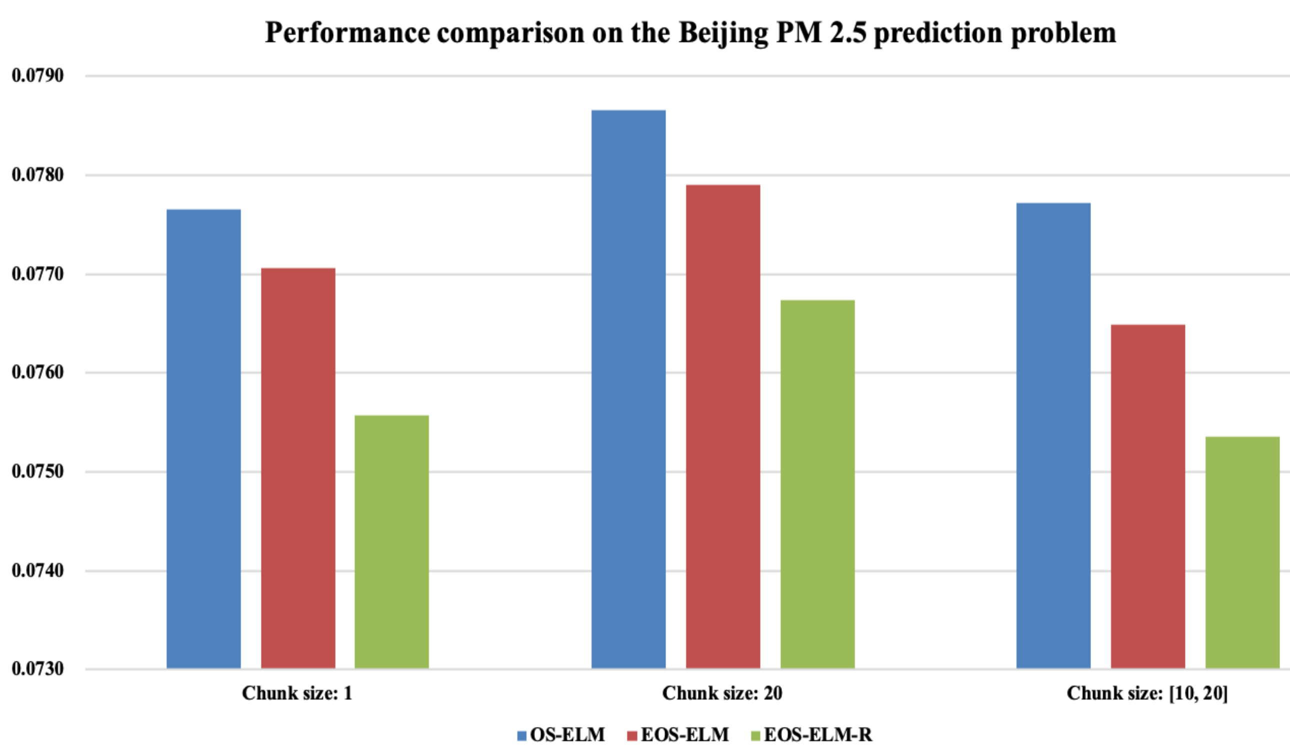

Fig. 3. Performance comparison of OS-ELM, EOS-ELM, and EOS-ELM-R on Beijing PM 2.5 prediction problem.

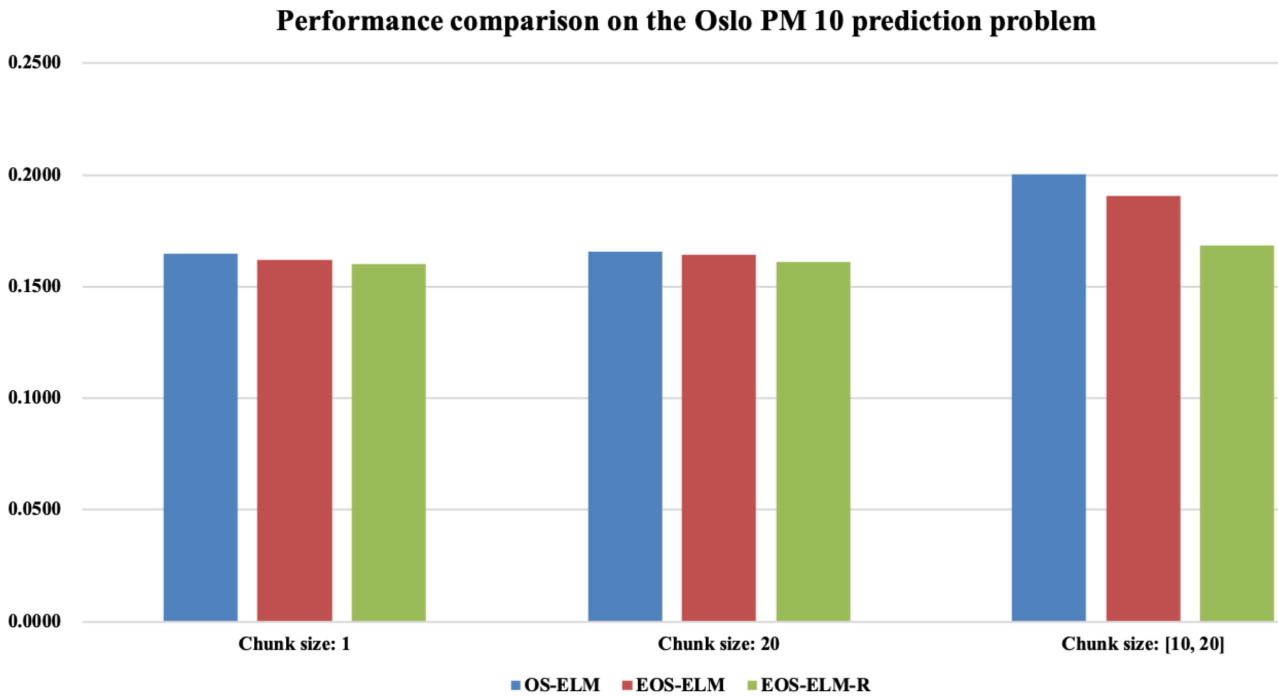

Fig. 4. Performance comparison of OS-ELM, EOS-ELM, and EOS-ELM-R on Oslo PM 10 prediction problem.

From Tables II-IV and Figs. 3-4, one can observe that EOS-ELM-R can achieve the smallest RMSE in all cases, which means that it has better generalization performance than OS-ELM and EOS-ELM. And the experimental results imply that using multiple distributions to initialize the base OS-ELMs can effectively improve the stability of the model.

V. Conclusions

To provide an algorithm with high training efficiency and good generalization ability for air quality prediction, we propose a novel ensemble OS-ELM (i.e., EOS-ELM-R) in this paper. EOS-ELM-R uses different distribution functions to initialize the random parameters of base OS-ELMs, which effectively improves the diversity of the base models. The prediction result of the final model is determined by averaging the predictions of the base models. We have fully verified the effectiveness of the proposed algorithm on two air quality prediction problems collected from the real world. In the future, we will consider using the statistical characteristics of the data to further improve the diversity of the base models [15].

Acknowledgment

The authors would like to thank the conference committee and anonymous reviewers for their valuable reviews and suggestions.

References

[1] Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems. 2019 Jan 28;30(11):3212-3232.

[2] Chen J, Ran X. Deep Learning With Edge Computing: A Review. Proceedings of the IEEE. 2019 Aug 1;107(8):1655-1674.

[3] Kök I, Şimşek MU, Özdemir S. A deep learning model for air quality prediction in smart cities. In2017 IEEE International Conference on Big Data (Big Data) 2017 Dec 11 (pp. 1983-1990). IEEE.

[4] Wang J, Song G. A deep spatial-temporal ensemble model for air quality prediction. Neurocomputing. 2018 Nov 7;314:198-206.

[5] Yi X, Zhang J, Wang Z, Li T, Zheng Y. Deep distributed fusion network for air quality prediction. InProceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2018 Jul 19 (pp. 965-973).

[6] Gu K, Zhou Y, Sun H, Zhao L, Liu S. Prediction of air quality in Shenzhen based on neural network algorithm. Neural Computing and Applications. 2019 Oct 11:1-4.

[7] Kang GK, Gao JZ, Chiao S, Lu S, Xie G. Air quality prediction: Big data and machine learning approaches. International Journal of Environmental Science and Development. 2018 Jan;9(1):8-16.

[8] Cao W, Wang X, Ming Z, Gao J. A review on neural networks with random weights. Neurocomputing. 2018 Jan 31;275:278-287.

[9] Jiang F, He J, Tian T. A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction. Applied Soft Computing. 2019 Dec 1;85:105827.

[10] Liu B, Yan S, Li J, Li Y. Forecasting PM2. 5 concentration using spatio-temporal extreme learning machine. In2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) 2016 Dec 18 (pp. 950-953). IEEE.

[11] Liang NY, Huang GB, Saratchandran P, Sundararajan N. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Transactions on neural networks. 2006 Nov 13;17(6):1411-1423.

[12] Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX. Assessing Beijing's PM2. 5 pollution: severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2015 Oct 8;471(2182):20150257.

[13] Cao W, Xie Z, Li J, Xu Z, Ming Z, Wang X. Bidirectional stochastic configuration network for regression problems. Neural Networks. 2021 Aug 1;140:237-246.

[14] Lan Y, Soh YC, Huang GB. Ensemble of online sequential extreme learning machine. Neurocomputing. 2009 Aug 1;72(13-15):3391-5.

[15] Cao W, Patwary MJ, Yang P, Wang X, Ming Z. An initial study on the relationship between meta features of dataset and the initialization of nnrw. In2019 International Joint Conference on Neural Networks (IJCNN) 2019 Jul 14 (pp. 1-8). IEEE.