Prediction of Unburned Carbon in Flue Dust based on Teacher-student Ensemble Method

Kuan Lu1*, Xingsen Yang1, Haichao Wang1, Ke Liu1, Xuhui Zhang1, Xiangrong Meng1, and Jun Li1

1 State Grid Shandong Electric Power Research Institute, Jinan, Shandong, 250021, China

*Corresponding author’s e-mail: lk83@163.com

Abstract: Based on the idea of knowledge distillation, a teacher-student ensemble method is proposed for UCFD (Unburned Carbon in Flue Dust) prediction when influencing factors have different sampling frequency. First, a multi-layer feed forward network is built as the teacher model. Loss function is customized to account for the sampling period discrepancy between outputs and inputs. With this teacher model, UCFD could have the same rate with its influencing factors. Second, Xgboost and Adaboost are used to form an ensemble student model to improve the training process and prediction robustness. Third, a power plant in Shandong province is chosen to make data experiment. Results illustrate that the teacher-student ensemble method can give a more accurate forecast.

1. Introduction
The two main factors that affect boiler combustion thermal efficiency are exhaust heat loss and incomplete combustion heat loss [1-2]. The UCFD, as an important technical indicator of the working efficiency of coal-fired boilers helps to guide the optimization of the combustion of the boiler. At present, most domestic coal-fired power plants use traditional manual sampling methods to detect UCFD. Although it has the advantage of high measurement accuracy, the whole process is complicated and difficult to reflect the actual combustion conditions in time. Therefore, it is vital to propose a timely and accurate UCFD prediction method.

At present, soft-sensing method of UCFD [3] is mainly to collect auxiliary variables that are easy to detect and closely related to the unburned carbon. Statistical methods are applied to make estimation and prediction. Most of them are machine learning methods based on scenarios where the sampling frequency of the independent variable and the dependent variable are consistent. Fang Wang [4] et al. proposed a random forest method to do feature engineering then predict the UCFD. Chunlin Wang et al. [5-6] established a support vector machine model for predicting the unburned carbon content in the fly ash of a high capacity boiler. Rui Cui [7] et al. used BP (Back-Propagation) neural network to predict the carbon content of fly ash, and combined with LM (Levenberg-Marquardt) optimization algorithm for network training. Xinmu Zhao [8] et al. established the 11-23-1 type BP neural network for fly ash carbon prediction and performed a single factor influencing factor analysis. Xugang Feng., et al [9] proposed the use of genetic neural network as the carbon content prediction method.
In the actual operation of the power plant, as UCFD needs to be obtained through the LOI (Loss On Ignition) process, of which the period (hour-level) is much longer than the sampling period of power plant operating factors (minute-level). This means that there are limited effective data pairs for the supervised learning. Therefore, large amount of daily operating data cannot be effectively used. Based on the idea of knowledge distillation, a teacher-student [10-11] ensemble method is proposed for UCFD prediction when influencing factors have different sampling frequency. First, a multi-layer feed forward network is built as the teacher model. Loss function is customized to account for the sampling period discrepancy between outputs and inputs. With this teacher model, UCFD could have the same data rate with its influencing factors. Second, Xgboost and Adaboost are used to form an ensemble student model to improve the training process and prediction robustness. Third, a power plant in Shandong province is chosen to make data experiment. Results illustrate that the teacher-student ensemble method can give a more accurate forecast.

2. Influential factors and data pre-process

The factors affecting the UCFD are mainly divided into two categories: unit operating factors and coal quality factors. The operating factors mainly include the active power of the generator, the oxygen content of the imported SCR flue gas, the air volume of the primary and secondary fans, etc.; the coal quality factors mainly include the moisture, ash, and air volatile content of the air dryer, etc. Table 1 shows the variables used in the modelling process. Among them, the unit operating factors are sampled per minute, while coal quality factors are sampled every 6 hours. For the coal quality data, it is believed that it remains unchanged during every sampling period. Finally, taking account of the LOI, the UCFD can be obtained every 2 hours.

Table 1. Influencing factors.

| Unit operating factors a | Symbol   | Coal quality factors b | Symbol   |
|--------------------------|----------|------------------------|----------|
| Power                    | x_b1     | Total moisture         | x_a1     |
| A\B\C\D\E\F side Coal  | x_b2-7   | Air dry base moisture  | x_a2     |
| Grinding Motor Current   |          |                        |          |
| A\B side Air preheater inlet flue gas temperature | x_b8-9 | Air dry base ash      | x_a3     |
| A\B side Primary fan air volume | x_b10-11 | Air dry base volatile matter | x_a4 |
| A\B side Induced draft fan inlet Flue gas pressure | x_b12-13 | Air dry based total sulfur | x_a5 |
| A\B side SCR inlet flue gas oxygen content | x_b13-14 | Receive base low heat | x_a6 |
| A\B\C\D\E\F side Coal feeder | x_b15 | Fixed carbon            | x_a7     |
| Instantaneous Coal feed  |          |                        |          |
| Main steam pressure      | x_b17    |                        |          |
| Main steam temperature   | x_b18    |                        |          |
| Total air volume         | x_b19    |                        |          |
| A\B side Induced draft fan inlet flue gas pressure | x_b20-21 |                |
| Air volume of A/B Air preheater | x_b22-23 |                |
| A\B side Outlet secondary air temperature | x_b24-25 |                |
| NOX Concentration of flue gas at the outlet of Desulfurization tower | x_b26 |                |

| Outputs                | Symbol  |
|------------------------|---------|
| UCFD                   | y       |

a 25 operating factors.
b 7 coal quality factors.
We collected unit operating, coal quality as well as UCFD data of a power plant in Shandong Province for modelling. Because there’re missing values and outliers in the data, pre-processing work are shown in Figure 1 to improve data quality. Taking into account the similarity among some factors, such as the current of six different mill motors, the amount of coal fed by six coal feeders, etc., dimension of the inputs can be reduced from 32 to 22.

3. Teacher-student ensemble framework

3.1 Teacher Model
In the field of machine learning and cognitive science, ANN (Artificial Neural Network), is a mathematical model that imitates the structure and function of biological neural networks. Working under the connection of a large number of artificial neurons, ANN is usually used to estimate or approximate a function. The teacher model uses a multilayer feed forward neural network to learn the mapping from independent variables to dependent variables, shown in the following figure:

3.2 Student Model
The Student model uses two machine learning algorithms to make a further training. Here, we integrated Xgboost [12] and AdaBoost [13] as an ensemble. The Xgboost algorithm trains a tree in each round, so that the loss function can be minimized. Its objective function not only measures the fitting error of the model, but also adds a regularization term, that is, the complexity of each tree. This penalty term plays the role in limiting the complexity of the tree and prevents over-fitting. The loss function is as follows:
\[ L = \sum_{i=1}^{n} I(\tilde{y}_i, y_i) + \gamma T + \frac{\lambda}{2} ||W||^2 \]

Pseudo code for the algorithm is as follows:

Table 2. Xgboost algorithm.

```
Xgboost
Initialize \( f_0(x) = \text{arg min}_c \sum_{i=1}^{n} L(y_i) \)

For \( t = 1, 2, \ldots, T \)

(1.1) For \( j = 1, 2, \ldots, J \), do: \( f_j^t = \text{arg min}_c \sum_{x \in \mathcal{R}_j} L(y_i, h_{t-1}(x)) \)

(1.2) Update: \( h_t(x) = h_{t-1}(x) + \sum_{j=1}^{J} f_j^t I(x \in \mathcal{R}_j) \)

(1.3) Final strong learner: \( H(x) = H_T(x) = \sum_{t=1}^{T} \sum_{j=1}^{J} f_j^t I(x \in \mathcal{R}_j) \)
```

AdaBoost refers to a particular method of training a boosted classifier. It can combine weak learners into a weighted sum that represent the final output of the boosted model. AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. Here are the algorithm:

Table 3. Adaboost algorithm.

```
Adaboost
Initialize sample weight \( w^1 \): \( w_i^1 = \frac{1}{n}, i = 1, 2, \ldots, n \)

For \( t = 1, 2, \ldots, T \)

(2.1) Use train set D and weights \( w^t \) to train a learner \( h_t(x) \)

(2.2) Calculate the maximum error: \( E_t = \max |y_i - h_t(x_i)| \)

(2.3) Calculate the relative error of each sample: \( e_i^t = \frac{y_i - h_t(x_i)}{E_t} \)

(2.4) Calculate the regression error rate: \( e_t = \sum_{i=1}^{n} w_i^t e_i^t \)

(2.5) Calculate the weight of the learner \( h_t(x) \): \( \alpha_t = \frac{e_t}{1 - e_t} \)

(2.6) Update \( w^{t+1} \): \( w_i^{t+1} \leftarrow \frac{w_i^t \alpha_t e_i^t}{Z_t} \), where \( Z_t \) is the normalization factor. \( Z_t = \sum_{i=1}^{n} w_i^t \alpha_t e_i^t \)

(2.7) Final strong learner: \( H(x) = \sum_{t=1}^{T} \ln \left( \frac{1}{\alpha_t} \right) h_t(x) \)
```

3.3 Teacher Student-Ensemble Framework

Inspired by knowledge distillation, we use the teacher-student ensemble method to improve the training process and prediction robustness. The teacher model is used to learn the mapping from high-frequency independent variables to low-frequency dependent variables to obtain a minute-level forecast output of UCFD. In the teacher training stage, a 4-layer neural network is selected. The network parameter structure is \((22, 512, 256, 1)\), where the input layer has 22 units and the hidden layers has 512 and 256 units respectively. The customized loss function is as follows:

\[ L_{TSR-T} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| \sum_{j=1}^{t} \tilde{y}_i^T - y_i \right|}{|y_i|} \]

where \( \tilde{y}_i^T \) is the minute-level output of teacher model and \( y_i \) is the true UCFD; \( t \) is UCFD’s sampling period, here is 2 hours.

The student model is utilized to make a further training to the minute-level input-output pair from the teacher model. Its customized loss function is:

\[ L_{TSR-S} = \frac{1}{nt} \sum_{i=1}^{n} \sum_{j=1}^{t} \frac{\left| \tilde{y}_i^T - \tilde{y}_i^S \right|}{|\tilde{y}_i^T|} \]
Where, \( \hat{y}^T_{ij} \) is the output from the trained teacher model and \( \hat{y}^S_{ij} \) is the output from student model. In order to improve the prediction robustness, we use both Xgboost and Adaboost to form an ensemble student model in Fig 3.

![Teacher-student ensemble framework](image)

**Figure 3. Teacher-student ensemble framework.**

4. **Experiment and results**

To verify the effect of the method, we have selected two schemes of using the teacher model alone and the teacher-student ensemble model respectively. When using the two schemes for training, all the samples are shuffled. Then data is divided into training set and test set according to the ratio of 7:3. The prediction results are shown in table 4, which shows the average error of the model on the test set for 10 rounds. Standard deviation of the error is listed in the brackets. We can find that the test set error of the teacher-student ensemble method is reduced by 16.5% compared with the teacher model alone. In addition, the prediction accuracy of each data point from the latter method is more stable.

| Model                          | Testing error% |
|-------------------------------|----------------|
| Only teacher model            | 7.16(0.52)     |
| Teacher-student ensemble      | 5.98(0.13)     |

We compare the predicted results of the teacher-student ensemble method with the actual UCFD, as shown in Figure 4 below. It can be seen that this method can accurately predict the UCFD even when it fluctuates obviously.

![Teacher-student ensemble method prediction](image)

**Figure 4. Teacher-student ensemble method prediction.**
5. Conclusion
In this paper, we construct a teacher-student ensemble method in UCFD prediction. With the data experiment in a power plant of Shandong province, the method’s performance is verified. It shows that both the robustness and accuracy can be improved under an ensemble framework.

References
[1] Haisheng Jiang. (2013) Reason analysis and reducing method of high carbon content in fly ash. Heilongjiang Science and Technology Information., 28:118.
[2] Xiuli Shi. (2020) Analysis on combustion adjustment test for reducing carbon content of fly ash in a 145 mw circulating fluidized bed boiler. Shandong Electric Power., 47:77-80
[3] Xue Li. (2018) Analysis and modeling for carbon content in fly ash based on field data. North China Electric Power University.
[4] Fang Wang, Suxia Ma, Ke Wang. (2018) Prediction model of carbon content in fly ash using random forest variable selection method. Thermal Power Generation., 11: 89-95.
[5] Chunlin Wang, Hao Zhou, Zhanghua Zhou, et al. (2005) Support vector machine modeling on the unburned carbon in fly ash. Proceedings of the CSEE., 20:72-76.
[6] Lin Li. (2013) Research on softer sensor method for unburned carbon in fly ash based on support vector machine. North China Electric Power University.
[7] Rui Cui, Guodong Liu, Xiaojiang Li. (2016) Measurement system for unburned carbon in fly ash based on l-m algorithm optimized BP network. Shanxi Electric Power., 3: 49-51.
[8] Xinmu Zhao, Chengliang Wang, Junfu Lv, et al. (2005) The investigation of carbon content in fly ash for a BP neural network-based pulverized coal-fired boiler. Journal of Engineering for Thermal Energy and Power., 02:158-162.
[9] Xugang Feng, Jiajun Qian, Jiayan Zhang. (2016) Prediction method of unburned carbon content in fly ash based on genetic neural network with sensitivity analysis. Journal of Electronic Measurement and Instrumentation., 07:1083-1089.
[10] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. (2015) Distilling the knowledge in a neural network. NIPS. Montreal, Quebec, Canada., pp.1–9.
[11] Muhamad Risqi U. Saputra, Pedro P. B. de Gusmao, Yasin Almalioglu, et al. (2019) Distilling knowledge from a deep pose regressor network. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 263-272
[12] Tianqi Chen and Carlos Guestrin. (2016) XGBoost: a scalable tree boosting system. In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016., pp. 785-794.
[13] Zhihua Zhou. (2016) Machine Learning. Tsinghua University Press Publishing, Beijing.