Citation:
W. Chen, C. Zhai, X. Wang, J. Li, P. Lv and C. Liu, "GCN and GRU Based Intelligent Model for Temperature Prediction of Local Heating Surfaces," in IEEE Transactions on Industrial Informatics, 2022, doi: 10.1109/TII.2022.3193414.

DOI:
https://doi.org/10.1109/TII.2022.3193414

GCN and GRU based Intelligent Model for Temperature Prediction of Local Heating Surfaces

Wanghu Chen, Chenhan Zhai, Xin Wang, Jing Li, Pengbo Lv and Chen Liu

***Abstract*— A boiler heating surface is composed of hundreds of tubes, whose temperatures may be different because of their positions, the influences of attempering water and flue gas. Using a criteria based on DBI, we propose to partition a heating surface into local ones, whose interactions in temperature are represented as a weighted Heating Surface Graph (HSG) at each point of time, and their current features are embedded in the HSG's nodes. Then, a local heating surface temperature prediction model WGCN-GRU is proposed. Graph Convolutional Network (GCNs) receive a series of HSGs, and extract the features of local heating surfaces and their spatial dependences in a time window. Features output by GCNs are finally directed to Gated Recurrent Units (GRUs) for temperature predictions. Experiments show that WGCN-GRU can averagely maintain the prediction error below 0.5℃. Compared with other models, it can reduce the errors by a rate from 5.6% to 46.8%, and shows advantages in RMSE and R^2. It also shows that the node-to-node weights for GCN can reduce the prediction error by 11.4%.**

***Index Terms*— Heating Surface, Temperature Prediction, Spatial-Temporal Features, Graph Convolutional Network (GCN), Gated Recurrent Unit (GRU)**

I. Introduction

DUE to the increasing demands on electric power as well as its efficient, safe and clean production, intelligent technologies are expected to play more important roles in the production management of thermal power plants. As to thermal power production, boilers are indispensable for providing heat energy, which is converted into mechanical energy in the steam turbine to drive the electricity generator.

As illustrated in Fig. 1, when a coal-fired supercritical boiler begins to work, pulverized coal is pumped into the furnace with air [1]. As the burning of pulverized coal in the furnace, Water Cooling Wall keeps absorbing the radiant heat. Saturated steam produced in the wall with water evaporation will pass through the tubes of a series of superheaters in succession. During the process, the high temperature flue gas produced by the pulverized coal combustion keeps transferring heat to the tube surfaces of these superheaters, and the steam transported in their tubes will be heated further. When leaving from High Temperature (HT) Superheater, the high temperature steam will be superheated and transferred to High Pressure Turbine (HPT) for electricity production. To improve the efficiency and emission reduction, after being reheated by Low Temperature Reheater (LTR) and High Temperature Reheater (HTR), the low temperature and pressure steam left in HPT will be transferred to Intermedia Pressure Turbines (IPT) again.

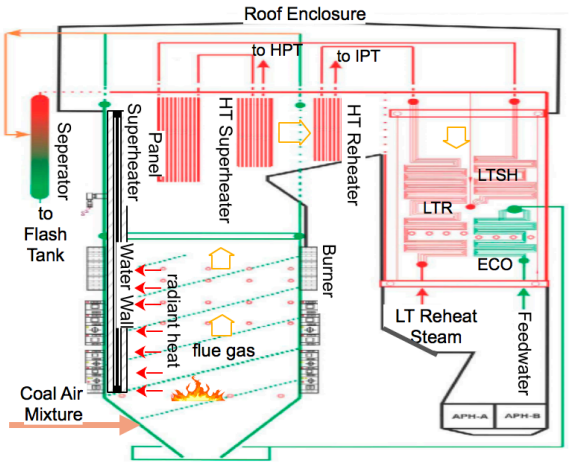


Fig. 1: Composition and working principles of a 600MW supercritical boiler.

Both superheater and reheater are also named heating surface because they absorb the radiant heat of combustion and the conduction heat of flue gas. During the production, a reasonable temperature control of the heating surface is critical, especially that of High Temperature Superheater, which supplies superheated steam for HPT. On the one hand, if the heating surface temperature is too high, tubes and inlets of

Manuscript received January 23, 2022. This work is supported by National Natural Science Foundation of China under Grant 61967013 and 61832004.

W. Chen is with the College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China (e-mail: chenwh@nwnu.edu.cn)

C. Zhai is with the College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China (e-mail: smilezchhhh@163.com)

X. Wang is with the Department of Information Systems, University of Maryland, Baltimore County, MD 21250, USA (e-mail: xinwang11@umbc.edu)

J. Li is with the College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China (e-mail: nwnulijing@nwnu.edu.cn).

P. Lv is with the College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China (e-mail: 925387739@qq.com).

C. Liu is with the School of Information Science and Technology, North China University of Technology, Beijing 100144, China (e-mail: liuchen@ncut.edu.cn)

HPT may be damaged. In the worst case, this damage may cause big disasters due to tube bursts. On the other hand, too low temperature can greatly affect the efficiency of electricity production and increase harmful gas emission [2]. So, it is important to maintain a high temperature as far as possible for a heating surface but avoiding overheating. During production, the adjustment of attempering water spray (Fig. 2) and flue gas flow plays an important role in temperature control of the heating surface.

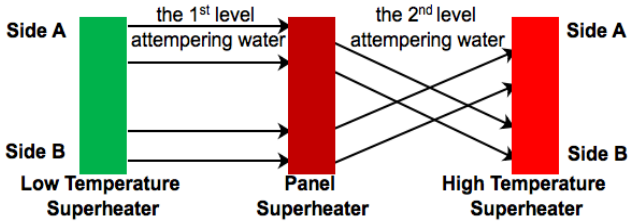


Fig. 2: Attempering water flows among various superheaters in a boiler.

The prediction of heating surface temperature is critical for the control systems to perform a better adjustment in attempering water, flue gas and so on [2]. So, model predictive control of heating surface temperature has drawn wide attention in recent years [3]. The parameters of a working boiler, such as tube temperature, attempering water spray and flue gas flow, can be gathered by various sensors [1], and are recorded as multivariate time series. Therefore, the paper intends to explore a data-driven approach to predict the temperature of heating surfaces depending on multivariate time series.

However, a superheater usually contains hundreds of tubes arranged as platens. These tubes often have different real-time temperatures because of the delayed effects of attempering water and the uneven temperature field of flue gas in the boiler. What's more, as time going, tubes in various inner spaces of a boiler may be in different conditions owing to their erosion, oxidation and so on. Assuming it is time-consuming to predict the temperature of hundreds of tubes in one heating surface, we can partition the heating surface into local ones and predict the temperature of each local heating surface. From the other aspect, we found that, when the temperature of one local heating surface rises or drops, so do some of others. We call such phenomena as spatial dependencies, which can be exploited to improve the temperature prediction. Assuming the average temperature of a heating surface keeps in a reasonable range, however, some of its tubes may become overheating so that disasters still could be caused. On the contrary, when the temperature of one of its tubes is too low, it may have been damaged. So, in the paper, the highest, mean and lowest temperature of all tubes in a local heating surface will be predicted. To be simple, we call them the highest, mean and lowest temperature of a local heating surface.

Therefore, this paper focuses on the prediction of local heating surface temperature by combining temporal features of multivariate time series and spatial dependency features among various local heating surfaces. It is well known that the Graph Convolutional Network (GCN) [4] is widely used in graph learning. It uses a graph convolution operation to obtain node embeddings by gathering the embeddings of its neighbours. In reality, the relations among local heating surfaces can be represented as a graph, which is called Heating Surface Graph (HSG) in the paper. We consider the partition of a heating surface as its tube clustering, and the number of nodes of a HSG will be determined based on a policy to minimize Davies-Bouldin Index (DBI), which is widely used to evaluate the effect of clustering. Besides, a Gated Recurrent Unit (GRU) is also be verified to have advantages in learning long-term dependencies in time series with its reset and update gate. Compared with Long Short-Term Memory (LSTM), a GRU has simpler structure and fewer parameters. Based on these, we explore the temperature prediction approaches for local heating surfaces based on GCN and GRU in this paper.

The expected contributions include: (1) A DBI based policy is proposed to partition a heating surface into local ones, whose interactions of temperature varying are represented with a HSG in each time window and features are embedded into each HSG node. (2) By receiving a series of HSGs, weighted and directed GCNs are proposed to extract the features of local heating surfaces and their spatial dependences in a time window, and their output features are delivered to GRUs for temperature predictions. (3) The proposed model WGCN-GRU is applied in real-world scenarios to predict the temperature of local heating surfaces of thermal boilers based on history temperature, attempering water, flue gases and so on. Experiments show that, compared with traditional models, the proposed model has great advantages.

II. Related work

Temperature control of heating surfaces is critical in the operation of boiler-turbine units. In past decades, various physics-based modelling methods, especially the computational fluid dynamics modelling methodology, are used to reveal the properties of superheated steam temperature [5]. A predictive controller is proposed in [2] to control the superheater steam temperature. For improving the superheater steam temperature control performance, a multi-objective optimization method is introduced in [6]. A generalized predictive control method based on a neuro-fuzzy network is developed for an ultra-supercritical power plant to improve control performance [7]. Physics-based methods can be used to explain the fundamental reasons for some phenomena. However, in practice, owing to sophisticated calculations with massive parameters adjustment and the non-linearity, high-order dynamics and disturbance variety of the control systems, physics-based methods have certain limitations in real-time predictions. So, in this paper, based on real-world scenarios, we focus on the data-driven temperature prediction of heating surfaces, which can help the control systems perform better.

Data-driven models can establish the nonlinear dynamic relationship between input and output variables directly by maximizing the conditional probability distribution, without the need for detailed mechanism knowledge [8]. As to the prediction of heating surface temperature, it is important to mimic the relationship between heating surface temperature

and its working conditions, such as attempering water, flue gas and so on. Usually, parameters about such working conditions of a boiler are recorded as multivariate time series. In many domains, multivariate time series have been applied to predict and infer real-world events based on machine learning models, including Support Vector Regression [9], Random Forest [10]. As a result, heating surface temperature prediction based on multivariate time series seems to be a feasible way. In [11], an Multi-Layer Perceptron (MLP) like ANN-based temperature prediction model for the steam temperature of the attemperator is explored. Although these approaches are verified to be effective, they have not given enough considerations to the temporal dependency of features which usually reflect the variation laws covered in the time series.

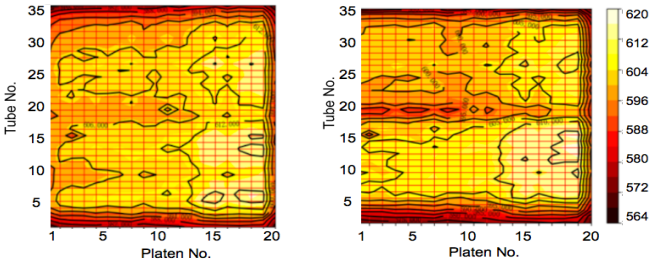
Recurrent neural networks (RNNs) are widely used for processing sequential data [12]. Its derived models, LSTM and GRU, can all avoid gradient vanishing and exploding in model training, which have good performance in learning long-term patterns [12], [13]. As the development in deep learning, LSTM and GRU are widely applied in time series prediction [14]. In [15], a model named T-LSTM is proposed to exploit the local information in weather time series prediction. Combining Empirical Model Decomposition with LSTM, a financial time series forecasting model is proposed in [16]. [17] proposed a novel hybrid model based on GRU and variational model decomposition for wind speed interval prediction. A recurrent neural network for water quality prediction called FM-GRU is presented in [18] based on a sequence-to-sequence (seq2seq) framework. In recent research, [19] proposed a data-driven ensemble model for the temperature prediction of superheated steam. A bidirectional GRU based approach with temporal self-attention mechanism is proposed to predict remaining useful life in the prognostics health management tasks [20]. In [21], a temperature forecasting approach for stored grain based on attention mechanism and LSTM is proposed. Its spatiotemporal feature extraction shows great advantages in improving forecasting accuracy. In [22], Residual Convolutional Neural Network (CNN) is used for the prediction of the material removal rate during the chemical mechanical polishing process. These studies can give us a solid foundation to explore the temperature prediction of heating surfaces based on multivariate time series.

Above all, current work mainly focuses on steam temperature prediction based on the inlet and outlet temperature of superheaters. To our best knowledge, there is little studies yet for data-driven heating surface temperature prediction depending on tube temperature, attempering water and flue gases. In addition, the paper intends to partition a heating surface into dozens of local ones based on tube clustering, and predict each one's temperature, which is very important in production management. Moreover, the potential laws in the temperature varying of various local heating surfaces under different work conditions can give knowledge for temperature prediction of local heating surfaces.

III. Approach to Temperature Prediction of Local Heating Surfaces

A. Architecture

With the view from overlook, tubes of a boiler superheater will be projected on a 2-dimensional plane. They are arranged into platens, each of which contains dozens of tubes. Thus, the real-time temperature distribution of a heating surface can be represented with a contour graph, whose horizontal and vertical axis refers to platen and tube numbers respectively. Fig. 3a illustrates the heating surface temperature distribution of High Temperature Superheater at the current time. In the figure, areas with different degrees of temperature are marked in different colors. The area with lighter color has a higher temperature. Fig. 3b shows the temperature distribution of the heating surface after 10 seconds. From Fig. 3, it is found that the temperature distribution on the heating surface shows some regular patterns. For example, the temperature on the top and bottom sides is lower than those of other areas. At the same time, the temperature distributions of nonadjacent areas may also have similarities. In fact, according to our studies, temperature variations of different areas often show some spatial dependencies. To describe the spatial dependencies of temperature variations in various areas, we partition a heating surface into local ones and summarise their relations in a graph.



(a) Temperature at the current time. (b) Temperature after 10 seconds.

Fig. 3: The real-time temperature contour of the superheater heating surface.

Therefore, we propose an approach to predict the temperature of local heating surfaces based on multivariate time series, whose architecture is shown in Fig. 4.

- Various sensors is used to monitor the real-time working and operation parameters of superheaters. The parameters include the temperature of each tube in a platen, the flow speed of first and second level attempering water on both sides of the heating surface, the flow speed of accidental attempering water and the flue gas. Because the heating surface locates in enclosed space with very high temperature, the extrinsic factors like weather have little influence to the work state of a thermal boiler.
- Data Acquisition System (DAS) gathers the parameters mentioned above every 10 seconds from sensors. The data gathered are preprocessed as multivariate time series. It is necessary to note that we will pay little attention to the data sensing and transferring in this paper.
- The heating surface is partitioned into local ones according to the temperature distribution of tubes.
- Heating Surface Graph is introduced to represent the features of each local heating surface and the relations

among all local heating surfaces. Thus, in any time window, there will be a series of Heating Surface Graphs.
- A GCN and GRU based model is constructed to predict the local heating surface temperature. It receives Heating Surface Graphs to achieve its training.
- The predicted temperature of each local heating surface is outputted to the boiler control system finally.

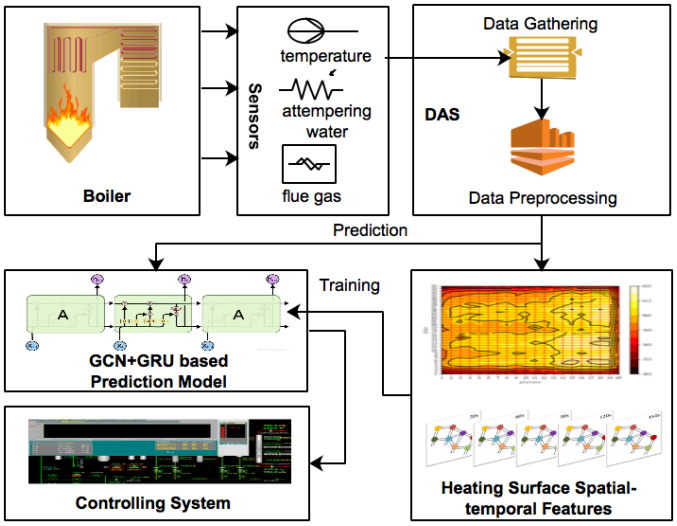


Fig. 4: Approach proposed to predict the temperature of local heating surfaces.

Partitioning local heating surfaces can be considered as the clustering of tubes in the heating surface. The objective is that the tubes in one cluster, also called the local heating surfaces, will show similarity in the temperature. It is well known that DBI is widely used to evaluate the clustering results and to obtain the optimal number of clusters. The smaller the DBI value, the better the clustering effect. Therefore, the number of local surfaces is determined based on the criterion of smallest DBI considering tubes in each local surface as a cluster.

As a result, the interactions among local heating surfaces in temperature varying can be represented as a graph named Heating Surface Graph (HSG). As time going, the working parameters of a local heating surface, such as attemperting water, accidental water and flue gas, may be very different. So, the weights on the edges of a HSG can help to reveal the impacts of its factors on local heating surface temperature. Compared with feature extraction techniques, it is easier to interpret the potential laws. At the same time, GRU aims to make each recurrent unit adaptively capture dependencies of different timescales, which is as powerful as LSTM even with small datasets. Although GRU and LSTM are ideal choices to achieve time series predictions, the former has much fewer parameters and simpler structure since each of its recurrent unit has only two gates and it does not need to separate memory cells. In fact, GRU is an improvement of LSTM as it also taking long-term dependencies into consideration with less computing power and training time. Therefore, we will explore the model for the temperature prediction of local heating surface based on GCN and GRU in this paper.

B. Heating Surface Graph Construction

Considering one heating surface as a 2-dimensional plane, it can be partitioned into multiple local heating surfaces, each of which refers to an area containing several tubes. The relations among all local heating surfaces can be represented in a graph. A Heating Surface Graph (HSG) at time t is defined as

$$G_{HS}^{(t)} = (S, R, W, X),$$

where S is a set of nodes, R is that of edges, W and X represent the weight and feature matrix respectively. In fact, each $S_i \in S$ refers to a local heating surface in an HSG. If there is $< S_i, S_j > \in R$, it means that the i^{th} local heating surface shows similarities in temperature variation with the j^{th} one, and $W_{i,j} \in W$ will specify such a similarity. Each row vector of X specifies the features of a local heating surface S_i, including its temperature, attempering water, accidental water and flue gas. Thus, the primary features of a heating surface in time window T can be represented as a series of HSGs. It is necessary to note that each feature of a local heating surface is aggregated from all tubes it contains.

For two local heating surfaces, the more the similarly of their temperature variation, the larger the weight of their relation. Given $S_i, S_j \in S, < S_i, S_j > \in R$, and their temperature variation vector V_i, V_j in current time window, the weight $w_{i,j} \in W$ is determined as following equation.

$$w_{i,j} = e^{\frac{-dist(V_i,V_j)^2}{\sigma_i^2}} \quad (1)$$

In Eq. 1, $dist(V_i, V_j)$ gives the Euclidean distance of two vectors V_i and V_j. Each element of V_i is the difference between the current element in temperature time series of the i^{th} local hearing surface and its directed predecessor. In the equation, σ_i means the standard deviation of all distances $\{dist(V_i, V_k)|k = 1, 2, ..., m\}$, where each S_k is a neighbor of S_i in the HSG. So, there will be $w_{i,j} \neq w_{j,i}$ usually when $i \neq j$. That is to say that HSG is directed. Either attempering water or flue gas can be considered flowing in a specific direction to some extent in the steam-water circulation system of a boiler. Thus, the temperature variations of two local heating surfaces may have different impacts on each other. Therefore, it is reasonable and necessary to define HSG as a directed graph. It is necessary to note that the local heating surface partition is a great challenge. In the paper, the local heating surface is partitioned into 2-dimensional grids, and there exist relations among all adjacent local heating surfaces.

Thus, given a local heating surface, its temperature can be predicted based on the temporal features extracted from related nodes in HSGs in a time window, as well as the spatial dependency features extracted from related nodes, which reflect the potential temperature variation dependencies among various heating surfaces.

IV. Local Heating Surface Temperature Prediction Model based on GCN and GRU

A. GCN and GRU based Prediction Model

According to the analytics above, for maintaining a higher temperature of the heating surface but avoiding overheating,

it is necessary to predict the temperature in advance. The proposed temperature prediction model of local heating surfaces is illustrated in Fig. 5. When a GCN receives a HSG at time t, the features of each local heating surface at time t, containing the spatial dependencies with its neighbours, can be obtained. So, GCNs are piled up to receive series of HSGs in each time windows, which are derived from the input multivariate time series. The output features of GCNs are then delivered to GRUs piled up to learn the temporal features of each local heating surface and finish its prediction its eventually.

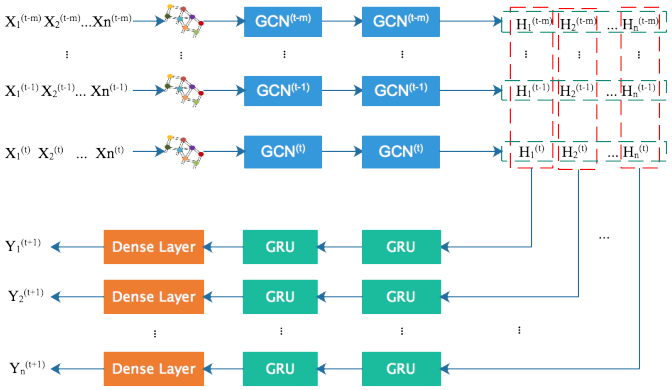


Fig. 5: The proposed temperature prediction model of heating surface.

As shown in Fig. 5, the model receives a feature matrix of all local heating surfaces as its input (Eq. 2), where $X_i^{(j)}$ represents the features of the i^{th} heating subsurface S_i at time j. The output of the proposed model is formulated as $(Y_1^{(t+1)}, Y_2^{(t+1)}, \cdots, Y_i^{(t+1)}, \cdots, Y_{n-1}^{(t+1)}, Y_n^{(t+1)})$, where $Y_i^{(t+1)} = (h_i, m_i, l_i)$, h_i, m_i and l_i is the highest, mean and lowest temperature of the i^{th} local heating surface at time $t+1$ respectively. Since the temperature of each local heating surface at time $t+1$ is predicted depending on its features from time $t-m$ to t, it means that the parameter $lookback = m$ in the prediction model.

$$X = \begin{pmatrix} X_1^{(t-m)} & X_2^{(t-m)} & \cdots X_{n-1}^{(t-m)} & X_n^{(t-m)} \\ \vdots & \vdots & \ddots & \vdots \\ X_1^{(t-1)} & X_2^{(t-1)} & \cdots X_{n-1}^{(t-1)} & X_n^{(t-1)} \\ X_1^{(t)} & X_2^{(t)} & \cdots X_{n-1}^{(t)} & X_n^{(t)} \end{pmatrix} \quad (2)$$

It is necessary to note that $X_i^{(j)}$=(HT, MT, LT, DSA, DSB, AWA, AWB, SFL) in this paper, where HT, MT, LT are the highest, mean, and lowest temperature of the tubes in the i^{th} heating subsurface S_i at time j; DSA and DAB give the flow speed of attempering water on side A and B of the heating surface respectively; AWA and AWB give the flow speed of accidental water on side A and B respectively; SFL is the flow speed of flue gas.

To learn the spatial dependencies among local heating surfaces, GCN is introduced to the proposed prediction model. After extracting spatial features using GCN, GRU is utilized for predicting the highest, mean and lowest temperature of each local heating surface at a specific time.

As shown in Fig. 5, m GCNs denoted as $GCN^{(j)}$ are piled up, where $j = (t-m, t-m+1, \cdots, t-1, t)$. It accepts heating surface temperature graph $HST^{(j)}$, which is generated from features of n local heating surfaces at time j, as its input. Spatial features of local heating surfaces at time j, $(H_1^{(j)}, H_2^{(j)}, \cdots, H_{n-1}^{(j)}, H_n^{(j)})$ will be then extracted by $GCN^{(j)}$. For improving the spatial feature extraction, several GCNs can be concatenated at each tile. Each GRU then accepts extracted features at one specific time for temporal prediction. Similarly, several GRUs can be concatenated to improve the effect. At last, the full connection layer will output the highest, mean and lowest temperature of local heating surfaces.

B. Spatial Feature Learning based on GCN for the Prediction

GCN is a kind of novel neural network, which can be used to learn graph-style data [23]. Traditional GCNs are usually built on undirected graphs, and can not assign various weights to the neighbors of a graph node. In a boiler, the temperature variation of two local heating surfaces may have different impacts on each other. So, the GCN model is extended to be directed and assigns weights to all graph edges based on HSGs. The GCN model proposed is defined as follows.

Given an HSG, $G_{HS} = (S, R, W, X, t)$, the GCN for time t used in the proposed model is defined as $f(X, A)$ logically, where $X = \{x_{i,j}\} \in R^{N \times F}$ $(1 \le i \le N)$ and $(1 \le j \le F)$ is a matrix of features of all graph nodes. Obviously, it satisfies $N = |S|$ and each graph node has F features. The adjacency matrix $A \in R^{N \times N}$ represents the relations among graph nodes. Given $a_{i,j} \in A$, if $a_{i,j} = 1$, there will exist a relation between local heating surface S_i and S_j in S, and an edge will also exist between two nodes in the HSG correspondingly. In the paper, because the heating surface are partitioned into grids, if there are $n = r \times c$ local heating surfaces that are arranged as r rows and c columns, the i^{th} local heating surface will have relations with the j^{th} one, where, $j = i-1, i+1, i-c, i+c$, and $(i-1) \mod c > 0$, $(i+1) \mod c > 0$, $(i-c) \mod r > 0$, and $(i+c) \mod r > 0$. In such situation, there is $a_{i,j} = 1$, where $a_{i,j} \in A$.

To considerate the self-connection of each graph node, we define an adjacency matrix of the heating surface graph as $\tilde{A} = A + I_N$, where I_N is an identity matrix [23]. Correspondingly, given a matrix D that represents the degree of each node of the graph, $\tilde{D}$ can be defined to involve self-connection of each node, where $\tilde{D}_{ii} = \sum_j A_{ij}$ and $\tilde{D}_{ij} = 0$ when $i \neq j$. Obviously, D and $\tilde{D}$ are all diagonal matrices, whose elements are all 0 except diagonal ones. We use the Hadamard product $\tilde{A} \cdot W$ to reflect the interactions among nodes on temperature varying, where W is a weight matrix computed with the help of HSGs according to Eq. 1. It is necessary to note that Hadamard product means each element in one matrix is multiplied by the element in another matrix at the same position. To avoid exploding gradients when there are too many nodes, the Laplacian transformation for symmetric normalization, $\tilde{D}^{-\frac{1}{2}} \tilde{A} \cdot W \tilde{D}^{-\frac{1}{2}}$, is conducted.

The proposed GCN model has several concatenated hidden layers. In each hidden layer, the feature of one node is

aggregated by its neighboring nodes in the previous layer. Supposing that the current hidden layer is the l^{th} one, the feature matrix of the $(l+1)^{th}$ layer, $H^{(l+1)}$, can be computed using the equation 3.

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A} \cdot W\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3}$$

In Eq. 3, $H^{(l)}$ is the feature matrix of the l^{th} layer, and $W^{(l)}$ represents the learnable weight matrix for neurons of the GCN. The input layer of the GCN model can be represented when $l = 0$. So, given a $GCN^{(t)}$ for spatial features learning at time t, there is $H^{(0)} = X^{(t)}$, where $X^{(t)}$ is a row vector in the feature matrix X (see in Eq. 2), which means the input features of all local heating surfaces at time t. By the concatenated GCNs at time t, the features of a node will be transformed to those of its neighbouring nodes. From Eq. 3, it can be found that the transformation depends on the degree matrix of each node, the adjacency relation matrix and the relation weight matrix denoted as $\tilde{D}$, $\tilde{A}$, and W respectively. Thus, the spatial features, which reflects the relations among local heating surfaces, are considered during the feature extractions of GCN nodes. In Eq. 3, σ is an activation function.

We assume that the proposed GCN model has N nodes and M edges, and each node is embedded with F features. Thus, in Eq. 3, there are $W \in R^{N\times N}$, $\tilde{A} \in R^{N\times N}$, $H^{(l)} \in R^{N\times F}$ and $W^{(l)} \in R^{F\times F}$. In fact, $M = \|A\|_0$ equals the number of nonzero values in the adjacency matrix $\tilde{A}$. The computation of the Hadamard product $\hat{A} = \tilde{A} \cdot W$ needs to cost $\mathcal{O}(N^2)$ for each layer beforehand. Supposing the model has L layers, it is verified that the total computation cost of the sparse-dense matrix multiplication $\tilde{D}^{-\frac{1}{2}}\hat{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}$ is $\mathcal{O}(LMF)$, and that of the feature transformation by applying $W^{(l)}$ is $O(LNF^2)$ [24]. Therefore, the computation complexity of the model in one gradient descendent will be $\mathcal{O}(LMF + LNF^2 + LN^2)$. In addition, the space complexity of the model is determined mainly by the storage of weights during the model training. So, the space complexity is $\mathcal{O}(LNF + LF^2)$.

C. Temporal Prediction based on Full-Connected GRU

Compared with LSTM, GRU has the advantage of adaptively capturing dependencies from different time scales. GRU is more efficient owing to its gating units that can modulate the information flowing inside the recurrent unit, without the separate memory cell and output gate [25], [26]. As a result, in the proposed model, we concatenate GRUs with GCNs to predict the temperature of local heating surfaces.

As shown in Fig. 5, the i^{th} layer of GRUs in the model receives a feature vector $(H_i^{(t-m)}...H_i^{(t-1)}H_i^{(t)})$, which is extracted by GCNs, to predict the temperature of i^{th} local heating surface at time $t+1$ depending on the spatial and temporal features from time $t-m$ to t. That is to say, for the i^{th} layer of GRUs, at time t, the recurrent unit receives elements of the vector $H_i^{(t)}$ as its input x_t. The prediction will be more effective since $H_i^{(t)}$ not only contains features of the i^{th} local heating surfaces related to temperature, attempering water, flue gas and accidental water, but also includes potential associations of its temperature variation with those of other local heating surfaces.

The output of the recurrent unit at time t is determined as following equations, where h_{t-1} is the previous activation, h_t is the candidate activation, z_t is an update gate which decides the update content, x_t is the input vector in the t^{th} time step, h_{t-1} holds the information for the previous $t-1$ units, and r_t is a reset gate which decides the reset content. The learnable weight matrices include W_z, W_r, W_h, U_z, U_r and U_h.

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{4}$$

where,

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ \tilde{h}_t &= tanh(W_h x_t + U_h(r_t \cdot h_{t-1})) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}) \end{aligned}$$

From the Eq. 4, it is found that the information from the hidden states at time $t-1$ will be forgotten by a rate of $1 - z_t$ and be remembered by a rate of z_t at time t. Then, based on the information from time $t-1$ along with the input features at time t, the output features at time t will be got. More importantly, the input features at time t for GRU were extracted by GCNs, which includes those related to historical temperature, attempering water, accidental water, flue gas as well as the dependencies among local heating surfaces. Thus, both the spatial and temporal feature of local heating surfaces will be extracted from the original time series. In addition, the dense layer concatenated to the GRU will transform the multidimensional features into one dimension. This improves the extraction of dependencies among various features.

The output of the GRU is computed finally with the equation $y_t = \sigma(h_t)$, and the training loss is computed with mean absolute errors. The fully-connected (FC) layer, also is well known as Dense layer, will transform the output of a GRU into a value that represents the temperature of one local heating surface at the next time.

In Eq. 4, supposing that the dimension of the hidden state is d_h and that of the input is d_i, there are $x_t \in R^{d_i}$, $h_{t-1} \in R^{d_h}$, and $W_r, W_z, W_h \in R^{d_i\times d_h}$ and $U_r, U_z, U_h \in R^{d_h\times d_h}$. It can be verified that $r_t, z_t \in R^{d_h}$. So, the total cost of computations for r_t, z_t and $\tilde{h}_t$ will be $3d_id_h + 3d_h^2$, and that for h_t will be $3d_id_h + 3d_h^2 + 2d_h^2 + d_h$ as a result. Therefore, when L GRUs are concatenated and a time series with the length T is given, the computation complexity of the GRU model in one gradient step can be denoted as $\mathcal{O}(LT(d_h^2 + d_hd_i + d_h))$. The space complexity of the GRU model is determined mainly by the storage of weights during the model training, which is $\mathcal{O}(LT(d_h^2 + d_hd_i))$.

V. Experiments and Analyses

A. Dataset Preparing

The dataset used in the experiments, which reflects the working conditions of a thermal boiler in 3 months, is provided by a power plant. Observations about the heating surface temperature, the flow speed of attempering water and flue gas are gathered every 10 seconds by various sensors. It is necessary to mentioned that the data is gathered by attaching sensors on all tubes in currently. It is supposing that, based

TABLE I: Illustration of observations contained in the dataset

Time	DSA-1	DSA-2	DSA-3	DSA-4	DSB-1	DSB-2	DSB-3	DSB-4
2/6/20 12:00:10 AM	39.35848	37.60057	9.740612	9.031951	18.50167	17.1874	15.187737	14.203365
2/6/20 12:00:20 AM	34.84987	32.70871	9.974368	10.01582	18.50167	17.34481	15.18774	14.06688
Time	SFL-1	SFL-2	SFL-3	SFL-4	AWA-1	AWA-2	AWB-1	AWB-2
2/6/20 12:00:10 AM	27.74641	27.9524	100.061	100.0229	5.16126	5.716673	0	0
2/6/20 12:00:20 AM	27.74641	27.9524	100.061	100.0229	7.317911	5.811957	3.815116	5.619773

TABLE II: The Pearson correlation coefficient between each independent variable X_i and each dependent variable Y_j

ρ_{X_i,Y_j} X_i / Y_j	DSA-1	DSA-2	DSA-3	DSA-4	DSB-1	DSB-2	DSB-3	DSB-4	SFL-1	SFL-2	SFL-3	SFL-4	AWA-1	AWA-2	AWA-3	AWA-4
Lowest	-0.35	-0.42	-0.51	-0.38	-0.28	-0.37	-0.65	-0.53	0.41	0.32	0.29	0.51	-0.35	-0.39	-0.21	-0.29
Mean	-0.39	-0.46	-0.49	-0.36	-0.52	-0.45	-0.61	-0.49	0.43	0.33	0.42	0.61	-0.42	-0.34	-0.38	-0.25
Highest	-0.52	-0.48	-0.29	-0.41	-0.57	-0.62	-0.34	-0.38	0.36	0.31	0.34	0.56	-0.52	-0.41	-0.54	-0.39

on the predicted temperature of local heating surfaces as well as the development of sensing technologies, there may exist a choice to attach sensors on each local heating surfaces. Thus, the data sensing in the boiler will be more efficient and economical. Table I represents some example observations contained in the dataset. In the table, DSA-X represents the speed of attempering water on side A of a superheater, and DSB-X represents those on side B. The column SFL-X represents the flow speed of flue gas, and AWA-X and AWB-X represent the flow speed of accident water on side A and B respectively. The dataset also contains the real-time temperature of each tube on the heating surface. The highest, mean and lowest temperature of each local heating surface is computed as important features for prediction model training. The 75% observations in the dataset are used for model training, and the remaining are used for model testing. The metrics for model evaluations include Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and the coefficient of determination (R^2).

If the input features of the model, including DSA-X, DSB-X, SFL-X, AWA-X and AWB-X, are considered as independent variables, the Highest, Mean and Lowest temperature to be predicted will be dependent variables. Table II represents the Pearson correlation coefficient ρ_{X_i,Y_j} between each independent variable X_i and dependent variable Y_j, where $1 \leq i \leq 16$ and $1 \leq j \leq 3$. It is found that, for each X_i, there is Y_j satisfying $|\rho_{X_i,Y_i}| > 0.3$. It means that each X_i moderately or strongly correlates with a dependent variable Y_j. We can also find that it satisfies $|\rho_{X_i,Y_j}| < 0.6$ for the majority of independent and dependent variables. This means that the relationships among independent and dependent variables are complex. So, the features used in the model are valuable for the highest, mean and lowest temperature prediction. Table II also shows that both attempering water and accidental water have negative correlation, but flue gases have positive correlation with the temperature of local heating surfaces.

The heating surface in the experiments has 700 tubes arranged as 20 platens. Thus, each platen has 35 tubes. Just mentioned in section III, the heating surface will be partitioned into grids. When the heating surface is partitioned into 28 local ones, each of which contains 5 tubes from 5 different platens. The minimum DBI equaling 0.186 will be got.

In experiments, the model parameter $lookback$ is set to 10 according to the prediction effects, which means that the temperature prediction of each local heating surface at time $t+1$ is depending on the time series from time $t-9$ to t. The original dataset used in the experiments is provided by sampling the working parameters of a supercritical boiler in every 10 seconds. In order to simulate a more practical evaluation, the original data is resampled with a step equals to 3 and 6 respectively. Thus, when a model is used to manage the production, the highest, mean and lowest temperature of each local heating surface after 30 or 60 seconds can be predicted based on its previous working parameters. It is usually feasible for the boiler control systems to give reasonable actions in 30 to 60 seconds according to the predicted temperature.

B. Evaluations of Prediction Based on Temporal Features

In this section, we first implement the FC-GRU based prediction model, then compare it with representative temporal prediction models, including MLP, RNN, FC-LSTM and Bi-LSTM. Table III and IV shows the evaluation results for the highest, mean and lowest temperature prediction of local heating surfaces 30 and 60 seconds later respectively.

Based on table III and IV, for the highest temperature prediction of local heating surfaces after 30 and 60 seconds, it is found that both MAEs of FC-GRU based model are much lower than 1°C, which equals 0.617 and 0.577 respectively. However, for the model of MLP, RNN and FC-LSTM, all their evaluation results are beyond or very close to 1°C. It is also found that Bi-LSTM has advantages in MAE compared with others except FC-GRU. Compared with MLP, RNN, FC-LSTM and Bi-LSTM, FC-GRU based model can reduce the MAE by 38%, 35%, 43% and 19% respectively when it is used to predict the highest temperature of local heating surfaces after 30 seconds. When to predict the highest temperature after 60 seconds, FC-GRU based model can also reduce the MAE by 32%, 34%, 29% and 24% respectively. Compared with MLP and RNN, the MAE of FC-LSTM decreases to 0.8661 when predicting the highest temperature after 60 seconds. However, for temperature after 30 seconds, the same factors can not generate rather well predictions. The reason may be that LSTM is general and effective when capturing long-term

TABLE III: Evaluation for the highest, mean and lowest temperature predictions of local heating surfaces after 30 seconds

Models	Highest Temperature					Mean Temperature					Lowest Temperature				
	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2
MLP	1.001	2.078	1.388	0.0016	0.9248	0.969	2.541	1.449	0.0016	0.9059	0.991	2.407	1.437	0.0016	0.9200
RNN	0.949	1.526	1.202	0.0015	0.9417	0.865	1.242	1.100	0.0014	0.9574	0.924	1.507	1.176	0.0015	0.9523
FC-LSTM	1.076	2.242	1.433	0.0018	0.9119	0.929	1.604	1.211	0.0015	0.9508	1.052	2.157	1.391	0.0018	0.9259
Bi-LSTM	0.736	0.971	0.963	0.0012	0.9669	0.638	0.737	0.817	0.0010	0.9778	0.701	0.871	0.911	0.0012	0.9709
FC-GRU	0.617	0.642	0.786	0.0010	0.9753	0.646	0.695	0.816	0.0010	0.9765	0.620	0.646	0.787	0.0010	0.9781

TABLE IV: Evaluation for the highest, mean and lowest temperature predictions of all 28 local heating surfaces after 60 seconds

Models	Highest Temperature					Mean Temperature					Lowest Temperature				
	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2
MLP	0.897	2.004	1.313	0.0014	0.9211	0.825	1.629	1.212	0.0013	0.9314	0.883	2.013	1.342	0.0015	0.9243
RNN	0.913	1.605	1.225	0.0015	0.9267	0.861	1.467	1.169	0.0014	0.9399	0.855	1.565	1.171	0.0014	0.9398
FC-LSTM	0.866	1.873	1.245	0.0014	0.9156	0.702	1.185	1.039	0.0011	0.9503	0.814	2.010	1.270	0.0013	0.9182
Bi-LSTM	0.713	1.054	0.993	0.0011	0.9578	0.622	0.841	0.872	0.001	0.9698	0.707	1.090	0.997	0.0012	0.96
FC-GRU	0.577	0.597	0.753	0.0010	0.9719	0.551	0.576	0.737	0.0009	0.9775	0.595	0.600	0.764	0.0010	0.9763

temporal dependencies [27]. Besides, FC-GRU based model also shows great advantages in MSE and RMSE compared with others. The MSE of FC-GRU based model is only 0.642, when all MSEs of the other models are larger than 1.5 except Bi-LSTM, whose MSE is also up to 0.971.

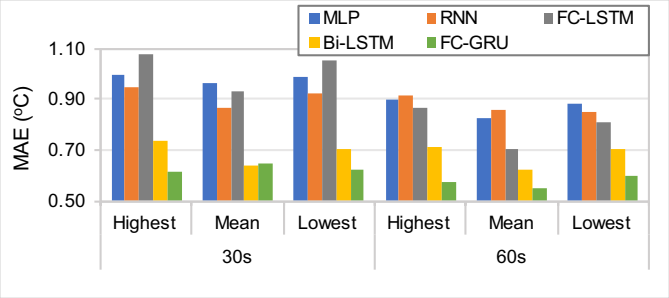


Fig. 6: MAEs comparison of various models for temperature prediction in local heating surfaces the 30s and 60s later.

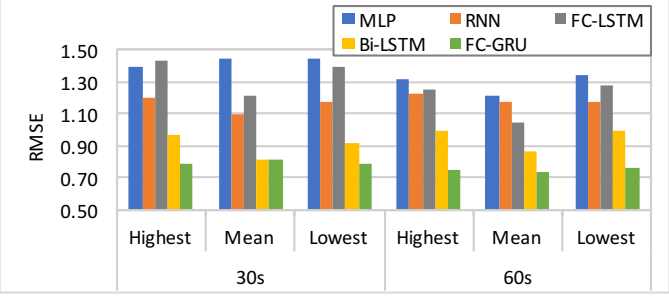


Fig. 7: RMSEs comparison of various models for temperature prediction in local heating surfaces the 30s and 60s later.

As to the mean and lowest temperature prediction of local heating surfaces, we can also find from table III and IV that FC-GRU has great advantages compared with other modes in both MAE and MSE. Fig. 6 and 7 illustrate the MAE and RMSE of each model respectively, which means FC-GRU based model achieves better accuracy and stability in temperature prediction. Moreover, Fig. 8 shows that the FC-GRU based model also has great advantages in R^2 index compared with others.

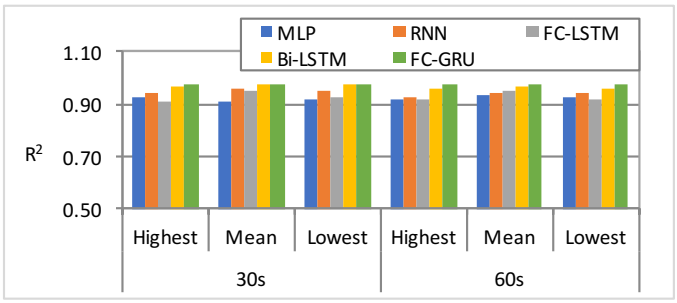


Fig. 8: R^2 indices comparison of various models for temperature prediction in local heating surfaces the 30s and 60s later.

The experiments above show that using FC-GRU to predict the temperature of the local heating surface depending on historical information of temperature, attempering water and flue gas is a good choice.

C. Evaluations to Prediction Based on Spatial-temporal Features

Based on the evaluation above, the proposed FC-GRU model only exploits temporal features of time series about temperature, attempering water and flue gas, which has shown good performances in temperature prediction of local heating surfaces. In order to improve the prediction performance, GCN is introduced to the spatial feature extraction among the local heating surfaces. We propose the model named GCN-GRU at first. GCN-GRU only considers whether there are relations among local heating surfaces, but it does not distinguish the intensities of these relations. We also evaluate WGCN-GRU in this section, which is an optimized model of GCN-GRU. WGCN-GRU determines the relation weights among local heating surfaces depending on their similarities of temperature variations (Eq. 1).

Fig. 9 and 10 illustrate the highest temperature prediction of a given local heating surface. The red curves represent the actual temperature and the blue ones represent the predicted value calculated from a specific model. The figure shows that, either predicting the temperature after 30 and 60 seconds, the model WGCN-GRU has the best performance. It is also found that GCN-GRU, Attention-LSTM and Attention-GRU

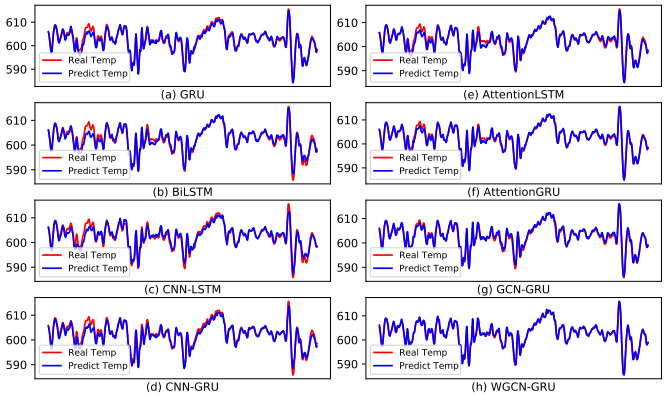


Fig. 9: The error comparison of the prediction of the highest temperature in a local heating surface the 30s later.

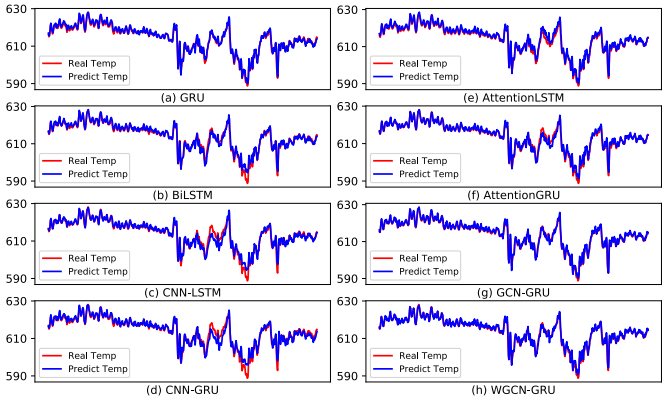


Fig. 10: The error comparison of the prediction of the highest temperature in a local heating surface the 60s later.

also show better performances compared with FC-GRU, Bi-LSTM, CNN-LSTM and CNN-GRU. This finding can verify our previous assumption that spatial features combined with temporal features of multiple time series can improve the prediction performance. The comparison between the WGCN-GRU and GCN-GRU based model turns out that the weighted relations are necessary and effective for the model prediction based on GCN and GRU.

Fig. 11 shows the test loss of WGCN-GRU and GCN-GRU. In the top sub-figure, the two models are trained to predict local heating surface temperature 30 seconds later, and the bottom one for that of 60 seconds later. It shows that WGCN-GRU has lower test loss. An interesting phenomenon is that LSTM does not show good performance as it is expected in time series prediction. It is valuable to be explored further in our future work.

We compare FC-GRU, CNN-LSTM, CNN-GRU, Attention-LSTM, Attention-GRU, GCN-GRU and WGCN-GRU in detail as follows. They are all used to predict the highest, mean and lowest temperature of local heating surfaces 30 and 60 seconds later respectively. The performances are shown in Table V and VI. As figure shown in Fig. 12 and 13, whatever to predict the highest, mean or lowest temperature of local heating surfaces after 30 or 60 seconds, WGCN-GRU achieves

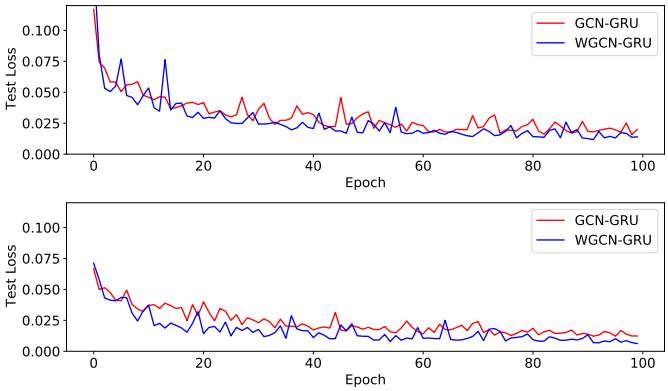


Fig. 11: Test loss of the model GCN-GRU and WGCN-GRU.

both the smallest MAEs and RSMEs compared with all other models. We also find that Attention-LSTM and Attention-GRU show advantages in MAE and RSME compared with other models except WGCN-GRU, and Attention-GRU surpasses Attention-LSTM. This means the proposed prediction model WGCN-GRU has higher accuracy and stability. In addition, Fig. 14 also shows that WGCN-GRU gets advantages in R^2 index compared to the other models.

The average MAEs of all models in predicting the highest, mean and lowest temperature of local heating surfaces are compared and shown in Fig. 15. The bars represent the average MAEs of various models while the curves reflect their average MAEs' increase for the model WGCN-GRU. It shows that, compared with the other models, WGCN-GRU can reduce the MAE by a rate of 9% to 56% when to predict the temperature 30 seconds later, and by a rate of 2% to 39% when to predict the temperature 60 seconds later. Averagely, WGCN-GRU can reduce the MAE by a rate of 5.6% to 46.8% compared with others. Particularly, compared with GCN-GRU, WGCN-GRU can reduce the prediction error by 11.4% averagely owing to its weighted node-to-node relations.

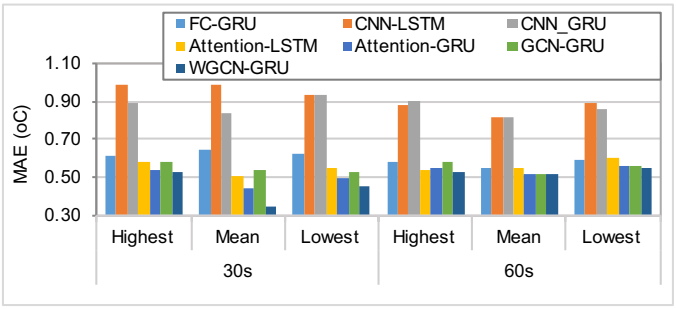


Fig. 12: MAEs comparison of various models for temperature prediction of local heating surfaces the 30s and 60s later.

Above all, depending on multivariate time series of history temperature, attempering water, flue gas and accidental water, the proposed model base on GCN and GRU is effective to predict the future temperature of local heating surfaces. The dependencies among local heating surfaces in temperature variations can be used to further improve temperature prediction. As a result, the proposed approach is valuable for the

TABLE V: Comparison of the prediction performances (30 seconds later)

Models	Highet Temperature					Mean Temperature					Lowest Temperature				
	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2
FC-GRU	0.617	0.642	0.786	0.0010	0.9753	0.646	0.695	0.816	0.0010	0.9765	0.620	0.646	0.787	0.0010	0.9781
CNN-LSTM	0.990	1.850	1.321	0.0016	0.9405	0.983	1.909	1.338	0.0015	0.9517	0.936	1.661	1.246	0.0016	0.9466
CNN-GRU	0.894	1.528	1.205	0.0014	0.9508	0.836	1.514	1.145	0.0014	0.9575	0.931	1.688	1.242	0.0015	0.9484
Attention-LSTM	0.587	0.630	0.770	0.0009	0.9789	0.510	0.451	0.656	0.0008	0.9848	0.552	0.577	0.730	0.0009	0.9813
Attention-GRU	0.536	0.513	0.688	0.0008	0.9832	0.446	0.349	0.573	0.0008	0.9857	0.492	0.447	0.627	0.0008	0.9850
GCN-GRU	0.587	0.594	0.753	0.0009	0.9801	0.542	0.481	0.683	0.0009	0.9836	0.525	0.464	0.669	0.0008	0.9842
WGCN-GRU	0.527	0.472	0.674	0.0008	0.9842	0.353	0.217	0.453	0.0008	0.9867	0.457	0.364	0.589	0.0007	0.9875

TABLE VI: Comparison of the prediction performances (60 seconds later)

Models	Highest Temperature					Mean Temperature					Lowest Temperature				
	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2	MAE	MSE	RMSE	MAPE	R^2
FC-GRU	0.577	0.597	0.753	0.0010	0.9719	0.551	0.576	0.737	0.0009	0.9775	0.595	0.600	0.764	0.0010	0.9763
CNN-LSTM	0.878	1.762	1.282	0.0014	0.9303	0.813	1.667	1.227	0.0013	0.9405	0.891	1.813	1.301	0.0015	0.9334
CNN-GRU	0.899	1.914	1.337	0.0014	0.9226	0.816	1.774	1.234	0.0013	0.9403	0.862	1.644	1.241	0.0014	0.9390
Attention-LSTM	0.535	0.538	0.725	0.0008	0.9761	0.548	0.652	0.762	0.0009	0.9764	0.601	0.787	0.850	0.0010	0.9701
Attention-GRU	0.547	0.569	0.739	0.0008	0.9761	0.513	0.531	0.699	0.0008	0.9802	0.563	0.630	0.760	0.0009	0.9756
GCN-GRU	0.579	0.587	0.751	0.0009	0.9801	0.514	0.458	0.661	0.0008	0.9817	0.559	0.563	0.724	0.0009	0.9799
WGCN-GRU	0.532	0.509	0.701	0.0008	0.9842	0.513	0.457	0.660	0.0008	0.9825	0.548	0.542	0.709	0.0009	0.9800

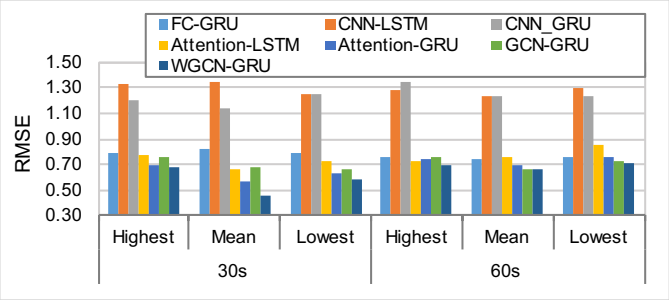


Fig. 13: RMSEs comparison of various models for temperature prediction of local heating surfaces the 30s and 60s later.

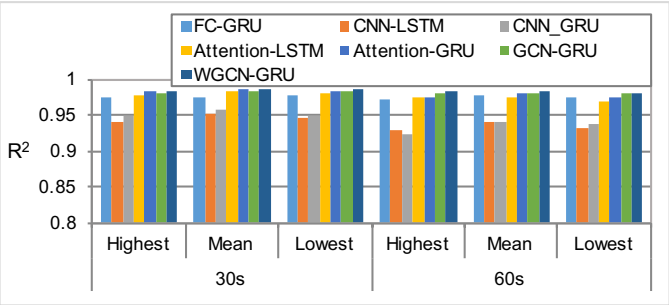


Fig. 14: R^2 indices comparison of various models for temperature prediction of local heating surfaces the 30s and 60s later.

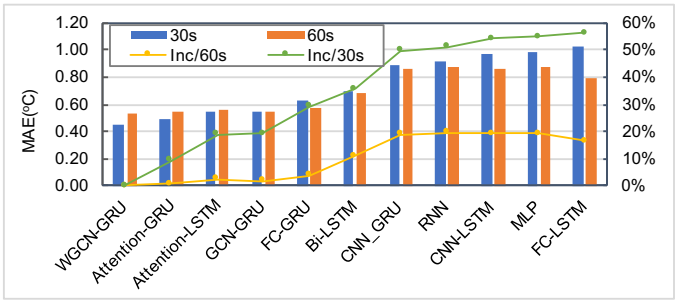


Fig. 15: Comprehensive comparison of various models' MAEs in predicting local heating surface temperatures the 30s and 60s later.

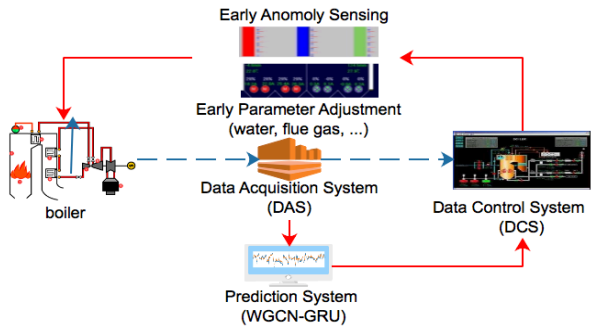


Fig. 16: The illustration to the applications of the proposed model in thermal power production.

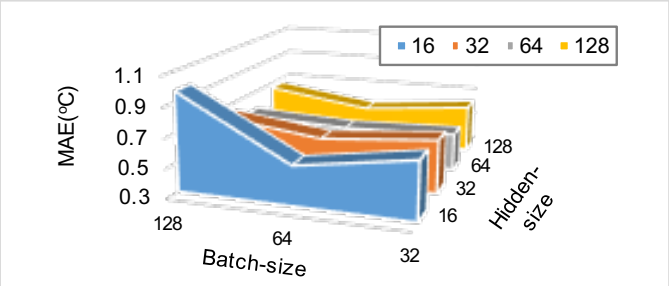
control systems to improve operational management.

As shown in Fig. 16, the proposed model can be used in the management of thermal power production. It keeps receiving the current working parameters from DAS, and predicting the future temperature of local heating surfaces. Depending on feedback from the prediction model, the DCS determines whether there are anomalies happening in the boiler. Moreover, the DCS can also adjust the working conditions, such as attempering water spay, flue gas flow and so on, according to the prediction results of the proposed model. This is very meaningful for safe and efficient production of thermal boilers.

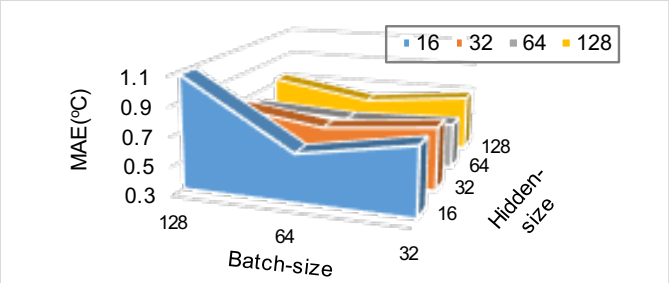
D. Discussions to the Model Complexity, Sensitivity and Interpretability

The proposed prediction model WGCN-GRU, has a layer for weight computation, 2 convolution layers for GCN, 2 GRU layers with 1 dropout and dense layer respectively. Because there are 28 local heating surfaces in the experiments, the total number of convolution layers are 2×28, and that for GRU layers are 2×28.

In order to interpret the effectivity of WGCN-GRU, its prediction MAEs are measured as the training Batch-size and the



(a) WGCN-GRU



(b) GCN-GRU

Fig. 17: The impacts of the Batch-size and Hidden-size on the model performances.

size of GRU Hidden Layers (Hidden-size) change. Moreover, its prediction MAEs in various conditions are also compared with those of GCN-GRU. It has been mentioned above that the proposed raw model GCN-GRU is similar to WGCN-GRU, but has no weights on node-to-node relations. Fig. 17a shows that WGCN-GRU performs better and more stable when the Batch-size equals 64, and gets the best performance when the Hidden-size equals 64 either. So does GCN-GRU as it is shown in Fig. 17b. Therefore, the Batch-size and the Hidden-size are both chosen as 64 finally. We can also find that, though the impacts of the Batch-size and Hidden-size on the prediction MAEs of the two models keep consistent, WGCN-GRU outperforms GCN-GRU in all conditions (Fig. 17). This verifies that WGCN-GRU is effective owing to the combination of weighted and directed GCN with GRU.

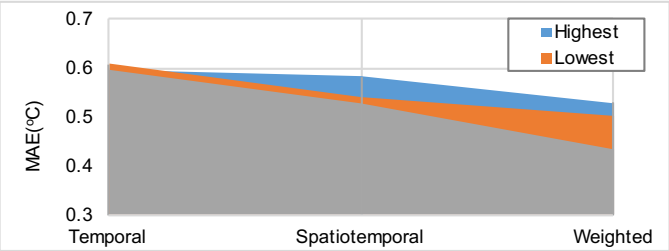


Fig. 18: Impacts of features on the performance of the proposed model.

In addition, to reflect the model sensitivity and interpretability, we use the Temporal, Spatiotemporal and Weighted spatiotemporal features respectively to predict the temperature of local heating surfaces. Fig. 18 shows that, no matter to predict the Highest, Mean or Lowest temperature, the spatial features contribute to the decline of prediction MAEs, and weighted spatiotemporal features can further reduce the prediction MAEs outstandingly. It means that the proposed model WGCN-GRU is reasonable and effective to extract spatial features with weighted GCNs to enhance the prediction.

Fig. 19 shows the data distributions of various kinds of features. To be convenient, features about attempering water, accidental water and flue gases are denoted as DT, AW and SFL respectively. HT, MT and LT represent the Highest, Mean and Lowest temperature of local heating surfaces respectively. It can be found that the data distributions of the features seem to be symmetric. At the same time, though there are lots of outliers in input features, WGCN-GRU still maintains a very small MAE, just as the evaluations in section V-C. So, WGCN-GRU is considered to be reliable and effective.

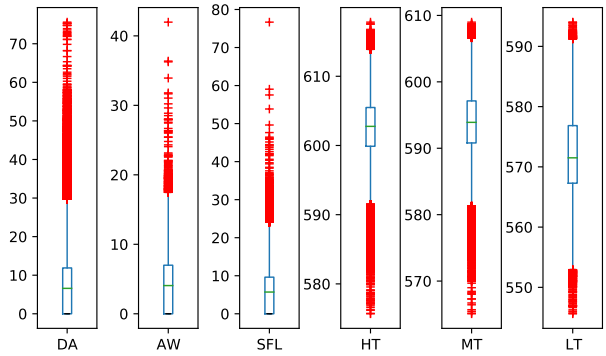


Fig. 19: The data distributions of the features used in the proposed model.

VI. Conclusion and Future Work

The paper proposed a GRU-based model for predicting the temperature of local heating surfaces in thermal boilers combined with GCNs. It is verified that depending on multivariate time series of tube temperature, the flow speed of attempering water and flue gas, the highest, mean and lowest temperature of a local heating surface can be predicted effectively. The proposed approach also provides an ideal choice for the prediction of multivariate time series based on both spatial and temporal features. Experiments with real-world datasets show that the proposed model has advantages over traditional models.

In this paper, the heating surface is partitioned into grids for constructing GCN models. In future work, we will explore irregular local heating surface partition approaches, based on clustering of multivariate time series, for optimising the representation of temperature interaction laws among local heating surfaces. Then, our temperature prediction model will be further improved based on such laws.

References

[1] W. M. Ashraf, G. M. Uddin, S. M. Arafat, S. Afghan, A. H. Kamal *et al.*, "Optimization of a 660 mwe supercritical power plant performance-a case of industry 4.0 in the data-driven operational management part 1. thermal efficiency," *Energies*, vol. 13, no. 21, p. 5592, 2020.

[2] X. Wu, J. Shen, Y. Li, and K. Y. Lee, "Fuzzy modeling and predictive control of superheater steam temperature for power plant," *ISA transactions*, vol. 56, pp. 241–251, 2015.

[3] Z. Wu, T. He, D. Li, Y. Xue, L. Sun, and L. Sun, "Superheated steam temperature control based on modified active disturbance rejection control," *Control Engineering Practice*, vol. 83, pp. 83–97, 2019.

[4] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2019.

[5] M. Granda, M. Trojan, and D. Taler, "Cfd analysis of steam superheater operation in steady and transient state," *Energy*, vol. 199, p. 117423, 2020.

[6] L. Sun, Q. Hua, J. Shen, Y. Xue, D. Li, and K. Y. Lee, "Multi-objective optimization for advanced superheater steam temperature control in a 300 mw power plant," *Applied energy*, vol. 208, pp. 592–606, 2017.

[7] C. Cheng, C. Peng, D. Zeng, Y. Gang, and H. Mi, "An improved neuro-fuzzy generalized predictive control of ultra-supercritical power plant," *Cognitive Computation*, pp. 1–8, 2021.

[8] A. Navarkar, V. R. Hasti, E. Deneke, and J. P. Gore, "A data-driven model for thermodynamic properties of a steam generator under cycling operation," *Energy*, vol. 211, p. 118973, 2020.

[9] C.H. Wu, J.M. Ho, and D.T. Lee, "Travel-time prediction with support vector regression," *IEEE transactions on intelligent transportation systems*, vol. 5, no. 4, pp. 276–281, 2004.

[10] D. S. Palmer, N. M. O'Boyle, R. C. Glen, and J. B. Mitchell, "Random forest models to predict aqueous solubility," *Journal of chemical information and modeling*, vol. 47, no. 1, pp. 150–158, 2007.

[11] J. Kim, H. Lee, J. Yu, J. Jang, J. Yoo, J. H. Park, and S. Kim, "Data-driven approach to attemperator steam temperature prediction in biomass power plant," *Journal of Electrical Engineering & Technology*, vol. 14, no. 4, pp. 1453–1462, 2019.

[12] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (indrnn): Building a longer and deeper rnn," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[13] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[14] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.

[15] Z. Karevan and J. A. Suykens, "Transductive lstm for time-series prediction: An application to weather forecasting," *Neural Networks*, vol. 125, pp. 1–9, 2020.

[16] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on ceemdan and lstm," *Physica A: Statistical Mechanics and its Applications*, vol. 519, pp. 127–139, 2019.

[17] C. Li, G. Tang, X. Xue, A. Saeed, and X. Hu, "Short-term wind speed interval prediction based on ensemble gru model," *IEEE transactions on sustainable energy*, vol. 11, no. 3, pp. 1370–1380, 2019.

[18] J. Xu, K. Wang, C. Lin, L. Xiao, X. Huang, and Y. Zhang, "Fm-gru: A time series prediction method for water quality based on seq2seq framework," *Water*, vol. 13, no. 8, p. 1031, 2021.

[19] P. Wang, F. Si, Y. Cao, Z. Shao, and S. Ren, "Prediction of superheated steam temperature for thermal power plants using a novel integrated method based on the hybrid model and attention mechanism," *Applied Thermal Engineering*, vol. 203, p. 117899, 2022.

[20] J. Zhang, Y. Jiang, S. Wu, X. Li, H. Luo, and S. Yin, "Prediction of remaining useful life based on bidirectional gated recurrent unit with temporal self-attention mechanism," *Reliability Engineering & System Safety*, p. 108297, 2022.

[21] S. Duan, W. Yang, X. Wang, S. Mao, and Y. Zhang, "Temperature forecasting for stored grain: A deep spatiotemporal attention approach," *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 17 147–17 160, 2021.

[22] J. Zhang, Y. Jiang, H. Luo, and S. Yin, "Prediction of material removal rate in chemical mechanical polishing via residual convolutional neural network," *Control Engineering Practice*, vol. 107, p. 104673, 2021.

[23] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.

[24] M. Chen, Z. Wei, B. Ding, Y. Li, Y. Yuan, X. Du, and J.-R. Wen, "Scalable graph neural networks via bidirectional propagation," *Advances in neural information processing systems*, vol. 33, pp. 14 556–14 566, 2020.

[25] C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, "Learned-norm pooling for deep feedforward and recurrent neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, pp. 530–546.

[26] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

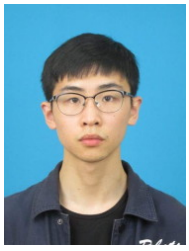
[27] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

Wanghu Chen is a Professor at the College of Computer Science and Engineering, Northwest Normal University, China. He got his Ph.D. degree in Computer Science from Institute of Computing Technology, Chinese Academy of Sciences in 2009. He was also a visiting scholar in University of California, San Diego from 2013 to 2014. His research interests include big data analytics, artificial intelligence and cloud computing.

Chenhan Zhai is working toward the M.S. degree at the College of Computer Sicence and Engineering, Northwest Normal University, China. His research interests include big data analytics, artificial intelligence and cloud computing.

Xin Wang is working toward the Ph.D. degree with the Big Data Analytics Lab in the Department of Information Systems, University of Maryland, Baltimore County. She is also working as a Research Assistant in the Center for Real-time Distributed Sensing and Autonomy (CARDS). Her research interests include distributed computing (systems), blockchains, big data analytics, federated learning, cloud computing and reproducibility.

Jing Li is an Associate Professor at the College of Computer Science and Engineering, Northwest Normal University, China. She got her Master degree in Computer Science from Lanzhou University, China in 2013. Her research interests include big data analytics, scientific workflow and service-oriented computing.



Pengbo LV is working toward the M.S. degree at the College of Computer Sicence and Engineering, Northwest Normal University, China. His research interests include big data analytics, artificial intelligence and cloud computing.

Chen Liu is a Associate Professor at the School of Information Science and Technology, North China University of Technology, China. He got his Ph.D. degree in Computer Science from Institute of Computing Technology, Chinese Academy of Sciences in 2008. His research interests include big data, artificial intelligence and service computing.