[[1]](#footnote-1)

Modified Clockwork Recurrent Neural Network for Multirate Industrial Soft Sensor

Shuchao Chang, Chunhui Zhao, *Senior* *Member, IEEE*, Yingnan Wang

***Abstract*—Multirate industrial scene, where the variables related to a process contains at least three kinds of measuring periods or sampling rates, is a common but troublesome phenomena for soft sensor modeling. Current methods generally use up-sampling or down-sampling techniques to transform the multirate data to a uniform sampling rate, which may bring deficiencies such as error accumulation, information loss, etc. In this paper, to overcome these drawbacks, the Modified ClockWork Recurrent Neural Network (MCW-RNN) is proposed for multirate industrial soft sensor. MCW-RNN first groups the hidden neurons corresponding to the sampling rates of process variables. Then at each moment, grouped hidden neurons are updated or remain unchanged according to the sampling results of their corresponding process variables. Such clockwork mechanism reasonably memorizes the information of process variables with different sampling rates in different grouped hidden neurons independently. At last, predictions for hard-to-measure variables can be calculated based on the updated hidden neurons. With the grouping for hidden neurons and the clockwork updating strategy, MCW-RNN adaptively and fully utilizes the multirate data, and elaborately explores the temporal relationships within different sampling rates. The effectiveness of MCW-RNN is demonstrated on a real coal mill case in power plant.**

***Index Terms*—multirate industrial processes, soft sensor, clockwork neural network.**

# I. Introduction

F

ast and accurate measurement of the critical process parameters is crucial in modern industrial processes. However, due to the harsh industrial environment, high sensor maintenance costs, and inadequate measurement technology, it is difficult to achieve real-time and reliable measurements of some key parameters using conventional measurement means [1]. Hence, the emergence of the soft sensor approach provides a viable means to do so.

Soft sensor technology achieves real-time estimation of hard-to-measure variables (referred to as quality variables) by constructing mathematical relationships between them and easy-to-measure variables (referred to as process variables) [2,3]. Thus, it is possible to obtain some important quality variables that cannot be directly measured by conventional sensors, which is important for the realization of operation status monitoring, and production process optimization.

Current research in the field of soft sensor is mainly focused on data-driven models [4]. Data-driven approach is based on historical operational data and uses statistical methods or artificial intelligence techniques to develop the models between process variables and quality variables. Because data-driven models do not require a detailed understanding of complex process mechanisms, they can greatly save the time and resources for modeling. At present, numerous multivariate statistical analysis methods [5,6] and machine learning methods [7-9] have been proposed for soft sensor tasks. To better handle the temporal relationships in most industrial processes, dynamic methods like Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), etc. have been applied to explore the temporal characteristics for soft sensor modeling [10,11].

Despite these aforementioned advances in soft sensor, in actual industrial scene, different variables have different importance and measurement difficulty, so they are generally measured by different measurement periods, or sampling rates [12,13]. For example, easy-to-measure process variables may be sampled at faster sampling rates such as seconds or minutes, while hard-to-measure quality variables may often be sampled once every few hours. Such phenomenon are common in practice and are generally named as multirate industrial scene. The multirate industrial scene is defined as when all the variables of the whole industrial process contain three or more kinds of sampling rates [14]. Currently, the research for multirate soft sensor generally first converts the multirate data to a uniform rate by up-sampling or down-sampling, and then the conventional modeling methods can be used. Up-sampling methods usually use interpolation or data imputation techniques to generate the missing values at the slower sampling rates, so the multirate data can be transformed to the fastest sampling rate [15,16]. On the contrary, down-sampling methods only use the complete samples with all the variables sampled, so the multirate data are transformed to the slowest sampling rate [17,18]. Nevertheless, such down-sampling operation will greatly waste the data, especially in the case when the scales of multiple sampling rates differ largely. Hence, data lifting technique provides a feasible solution which fuses all the incomplete samples within adjacent complete samples, so that the multirate data are transformed to the slowest sampling rate without information loss. Different from these conventional methods, Chai et al. [19] proposed a transfer learning based approach named Variational Progressive-Transfer Network (VPTN). Firstly, the multirate data are divided into different data chunks according to the number of sampled variables. Then, the Variational Auto-Encoder (VAE) models are sequentially transferred and constructed from the smallest chunk to the largest chunk, until the final model for soft sensing is obtained.

These aforementioned methods can address the multirate soft sensor modeling task to some extent, while some limitations still exist. The performance of the up-sampling methods highly depends on the imputation effect. When the missing rate is high, the error accumulation phenomena may greatly vulnerate the predictive performance. As for the down-sampling methods, although the data lifting technique can avoid the information loss, the temporal relationships at faster sampling rates will not be readily captured. Besides, the ratio of process variables at fast and slow sampling rates after data lifting operation will become imbalanced, which may overwhelm the information in process variables with slower sampling rates. While for VPTN, it is designed to effectively resolve the lack of labeled data, but the temporal relationships among samples are not considered. Moreover, down-sampling methods and VPTN can only predict quality variables at the slowest sampling rate, which will limit their application for real-time monitoring and control.

To handle the multirate soft sensor modeling more comprehensively, the ClockWork Recurrent Neural Network (CW-RNN [20]) gives us some enlightenment. CW-RNN is originally proposed to address the gradient exploding and vanishing problem by updating hidden neurons at different periods like a clockwork. It can be found that the process variables in multirate scene naturally possess different periods, i.e. sampling rates. Thus, the clockwork mechanism can adaptively update the hidden neurons according to the sampling results for process variables at each moment, which will be an appropriate pattern to handle the multirate data.

In this paper, a novel multirate industrial soft sensor, named MCW-RNN, is proposed. Inspired by the idea of clockwork mechanism, MCW-RNN effectively overcomes the error accumulation and information loss in conventional methods. Specifically, the hidden neurons are first grouped corresponding to the sampling rates of process variables. Then, the grouped hidden neurons are updated or remain unchanged group by group according to the sampling results for process variables at each moment. Such strategy can memorize the information of process variables at different sampling rates separately. Finally, the predictions for quality variables are calculated based on the hidden neurons after clockwork updating. MCW-RNN properly adapts to the multirate data structure with different process variables being treated equally, and elaborately extracts the temporal relationships within different sampling rates. Furthermore, MCW-RNN can give predictions at any sampling point, which will provide real-time measuring basis for process control and optimization. These properties all make MCW-RNN a simple but well-designed solution for the multirate soft sensor.

The main contributions of this paper are summarized as below:

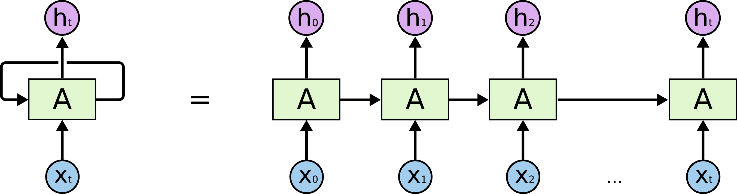
* The modified clockwork updating strategy proposed in this paper properly adapts to the multirate data, which makes process variables with different sampling rates fully and equally utilized.
* Through altering the sequence position, MCW-RNN can give predictions at any sampling point, which breaks the limitations of down-sampling methods and provides measuring basis for process control and optimization.

# II. Preliminaries

Some preliminary knowledge about recurrent neural network and clockwork recurrent neural network are introduced in this section.

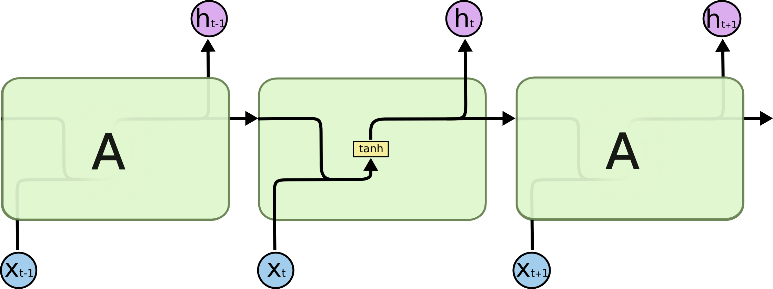
## A. Recurrent Neural Network

Recurrent Neural Network (RNN) [21] is a kind of neural network framework good at handling sequential data, which is implemented by its recurrent structure. Fig.1 is an illustration of the structure of RNN. and  in Fig.1 respectively denote the input and the output of hidden layer at time , and  represents the hidden layer. Compared with other conventional neural networks, in RNN, the output of hidden layer at time  is also taken as input at time . Such subtle design makes the information at past moments transmitted to the current moment, so the sequential relationship among sequential data can be effectively handled and captured by RNN. Due to the property of RNN, RNN has been widely applied in many tasks related to sequential data, e.g. natural language processing [22].



**Fig. 1.** An illustration of the structure of RNN [23].

The inner structure of the vanilla RNN is described by Fig.2 and formula (1).  and  denote the weights, and  is the bias. Hyperbolic tangent activation function, , is used here for nonlinearity. The computation formula of RNN is similar to that of Multi-Layer Perceptron (MLP), where the difference mainly lies in that the output of hidden layer at time  is involved.



**Fig. 2.** The inner structure of RNN [23].

 (1)

Despite the adaptability of RNN for handling sequential data, gradient vanishing [24] and gradient exploding [25] phenomena may occur when the sequence length is too long, which will severely affect the convergence performance. To address such problem, many improved versions of RNN have been proposed, such as Long Short-Term Memory (LSTM) [26], Gated Recurrent Unit (GRU) [27], etc., which can capture the short-term temporal relationship as well as the long-term.

## B. Clockwork Recurrent Neural Network

ClockWork Recurrent Neural Network (CW-RNN) is first proposed by Koutnik et.al. [20] at 2014. The ideology of CW-RNN is to adopt different updating periods for hidden neurons in order to settle the inapplicability for long-term sequential relationship of the vanilla RNN. At present, CW-RNN has already been applied in the field of traffic prediction [28], speech synthesis [29], etc. The computation rules of CW-RNN are illustrated in Fig.3.



**Fig. 3.** An illustration of the computation rules in CW-RNN [20].

From Fig.3, the hidden neurons are first endowed with different updating periods, i.e. “1, 2, 4, 8, 16”. When the time  is the integer multiple of an updating period, the corresponding hidden neuron will be updated. The time in Fig.3 is , so the hidden neurons with updating periods 1 and 2 will be updated, which are marked with shadow. Other hidden neurons will remain their states unchanged. Moreover, CW-RNN assumes that the hidden neurons with slower updating rates be able to influence the hidden neurons with faster updating rates while the faster neurons cannot conversely. Based on such assumption, the lower triangular block of the weight matrix  is set to zeros accordingly as shown, which can also reduce the amount of trainable parameters. The clockwork mechanism in CW-RNN can memorize the short-term and long-term temporal relationships in hidden neurons with fast and slow updating rates respectively, so that CW-RNN excellently implements the learning for short-term and long-term temporal relationship.

# III. Methodology

Our proposed MCW-RNN will be introduced in this section. The problem statement for multirate industrial soft sensor and the motivations for MCW-RNN are described first. Then, the clockwork updating strategy and the overall network structure of MCW-RNN are respectively presented in detail, followed by the specific application instruction.

## A. Problem Statement and Motivations

In this paper,  and  respectively denote the data matrices composed of process variables and quality variables, where ,  and  are the number of samples, process variables and quality variables respectively. In process industry,  and  are generally collected in the order of time, so  and  can present the process variables and quality variables of a sample collected at time .

In the multirate scene, it is assumed that there exist  kinds of sampling rates in total, where the process variables contain  kinds of sampling rates, and the sampling rates for different quality variables are uniform. In practice, if there exists multiple sampling rates for quality variables, it is readily to model for each kind of sampling rate individually. Then, the data matrix  can be permuted as  according to the order of the sampling rates, where  and . The  kinds of sampling rates are denoted as , where  is the basic sampling rate.­­­­

To give an intuitive cognition for the data structure in the multirate scene, an example with 4 kinds of sampling rates, i.e. , ,  and , is illustrated in Fig.4. The target of soft sensor in the multirate scene is to learn the function mapping between process variables and quality variables, i.e. , where  denotes the model parameters.



**Fig. 4.** An example with 4 kinds of sampling rates.

According to the description of multirate scene above, the process variables with different sampling rates can be seen as different updating rates. When the time satisfies the requirement of the specific sampling rate, the corresponding process variables will be measured and updated. It is not hard to find that such pattern is similar to the clockwork updating mechanism for hidden neurons in CW-RNN. The main difference lies in that the updating rates for hidden neurons in CW-RNN are set manually, while the process variables in multirate scene naturally possess different updating rates, i.e. sampling rates. Therefore, inspired by the updating mechanism in CW-RNN, when modeling soft sensor in the multirate scene, the hidden neurons can be grouped in accordance with the sampling rates of process variables. Each grouped hidden neurons adopts the sampling rate of its corresponding process variables as updating rate. The information of process variables with different sampling rates can be elaborately memorized in the corresponding hidden neurons, which will effectively overcome the error accumulation in up-sampling methods and information loss in down-sampling methods.

## B. Modified Clockwork Updating Strategy

Our proposed MCW-RNN modifies the clockwork mechanism in CW-RNN so that it can adapt to the multirate soft sensor scene. Fig.5 gives an illustration of the modified clockwork updating strategy in MCW-RNN based on the example in Fig.4. The process variables are grouped into , and , i.e. the green, blue and yellow blocks, according to the speed of the sampling rates. For simplicity, assume each group possess 2 process variables.



(a) When only , i.e. the green blocks, are sampled.



(b) When both  and , i.e. the green and blue blocks, are sampled.



(c) When ,  and , i.e. the green, blue and yellow blocks, are all sampled.

**Fig. 5.** The illustration for the modified clockwork updating strategy in MCW-RNN. The green, blue and yellow blocks respectively correspond to the grouped process variables with sampling rates ,  and  in Fig.4. The grey blocks are not used or updated.

First, the hidden neurons need to be grouped corresponding to the grouped process variables, which is formulated in formula (2) below.

 (2)

where  is the coefficient for dimension expansion. In the illustration in Fig.5, , so each grouped process variables are mapped from 2 dimensions to 3 dimensions in the hidden neurons. Because different process variable groups use the same coefficient , the information ratio among different groups will remain the same in the hidden neurons, which can make the process variables with different sampling rates treated equally in the following steps.

Then the hidden neurons can be updated group by group according to the sampling results for the process variables. In Fig.5(a), when only  are sampled, we only update the corresponding part in hidden neurons, i.e. . Because the dimension of  and  are 2 and 3, the first 3 rows in  and the left upper 3×2 parts in  are activated for calculation. When both  and  are sampled as shown in Fig.5(b), the corresponding parts in hidden neurons,  and , will be updated. Similarly, the first 6 rows in  and the left upper 6×4 parts in  are activated for calculation. Here, we do not assume that hidden neurons with slower updating rates affect the hidden neurons with faster updating rates unidirectionally as in CW-RNN. Instead, hidden neurons with different updating rates are treated equally and the mining for their correlations is beneficial for soft sensor modeling. Hence, the weight matrix  in MCW-RNN is differently designed that all the hidden neurons at the last moment can participate in the calculation for the current moment. When ,  and  are all sampled, all the hidden neurons will be updated, so all the blocks in  and  are activated, which is the same as the calculation form of the vanilla RNN.

The formula (3) below gives a mathematical description of the modified clockwork updating strategy in MCW-RNN.

 (3)

This formula is nearly the same as the calculation form of RNN in formula (1). The weight matrices  and  are expressed in the form of block matrix, where  and . When the time  satisfies the requirement of modulo division, the corresponding blocks will use the parameters of themselves, otherwise use the zero matrix. Formula (4) shows the rules for the values of  and .

 (4)

The modified clockwork updating strategy properly groups the hidden neurons corresponding to the sampling rates of the process variables, so that different grouped hidden neurons can independently update and memorize the process information under different sampling rates. While the proposed clockwork updating strategy only handles the mapping from process variables to hidden neurons, how to build the overall network structure of MCW-RNN for soft sensor modeling will be introduced in the next section.

## C. Overall Network Structure

Based on the introduction of the modified clockwork updating strategy, the overall network structure of MCW-RNN can be readily designed as shown in Fig.6. We continue to use the example in Fig.4 and Fig.5 for illustration. First, the initial hidden neurons  need to be initialized, generally the zero initialization. Then, the process variables are sequentially fed to update the hidden neurons with the modified clockwork updating strategy. Until the number of input samples reaches the sequence length (it is  in this example), the last hidden neurons  enter two MLP layers to give the predictions for quality variables. Moreover, unlike the down-sampling methods which can only predict the quality variables at the slowest sampling rate, MCW-RNN can give predictions at any sampling point by altering the sequence position.



**Fig. 6.** The overall network structure of MCW-RNN. The purple and orange arrows denote the direction of data flow. “MLP” is the multi-layer perceptron layer.

The computation formula of MLP is shown in formula (5).  and  respectively denote the input and output of MLP.  and  are the learnable weights and bias parameters in MLP. ReLU activation function, , is used in MLP for nonlinearity.

 (5)

The loss function of MCW-RNN is defined in formula (6), which is composed of two parts, i.e. the sum of prediction error and the regularization for model parameters.  and  respectively present the ground truth and the predictions for quality variables.  is the L2 norm, and  is the coefficient for the regularization term. All the learnable parameters in MCW-RNN are denoted as , which need to be optimized during training.

 (6)

## D. Application of MCW-RNN

In the offline training phase, gradient descent method [30] is applied to minimize the loss function defined in formula (6), and the optimal parameters can be obtained as shown in formula (7).

 (7)

In the online application phase, the process variables of the test samples are fed into the MCW-RNN with the optimal parameters  to obtain the predictions for quality variables, which is shown in formula (8).  and  respectively denote the process variables and the predictions of a test sample.

 (8)

With the modified clockwork updating strategy and the designed overall network structure, MCW-RNN properly connects the clockwork mechanism with the multirate soft sensor. Through corresponding grouping and updating, hidden neurons can memorize the information in process variables with different sampling rates separately, which avoids the dominance of process variables with faster sampling rates. MCW-RNN also elaborately explores the temporal relationships within multiple sampling rates. Furthermore, MCW-RNN can give predictions at any sampling point which provides reliable measuring basis for real-time control and optimization. Thus, MCW-RNN provides a simple but comprehensive solution for the multirate soft sensor task.

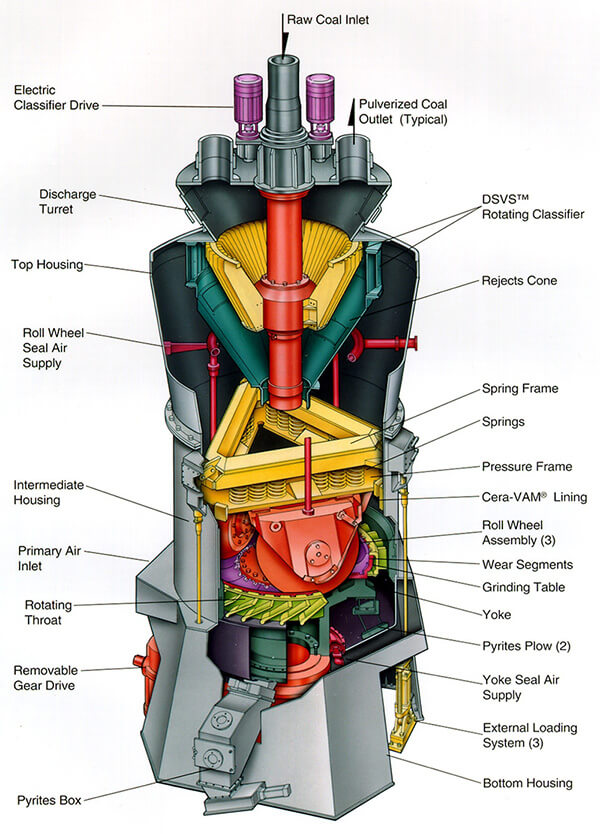
# IV. Case Study

To validate the effectiveness of the proposed MCW-RNN, a real coal mill case is experimented in this section. The coal mill equipment will be introduced first, followed by the detailed experiment setup and the comprehensive discussions about the results.

## A. Coal Mill Description

Coal mill is the preparation equipment of pulverized coal for boiler fuel in the power system and the professional equipment for preparing auxiliary materials for blast furnace ironmaking. The structure of a medium speed roller type coal mill is illustrated in Fig.7. Its grinding part is composed of rotating grinding ring and three fixed and rotatable grinding rollers rolling along the grinding ring. The raw coal to be grinded falls from the coal falling pipe in the center of the mill to the grinding ring, and the rotating grinding ring moves the raw coal to the grinding roller by means of centrifugal force, which is grinded by the grinding roller. The grinding force is generated by the hydraulic loading system. The grinding and drying of raw coal are carried out at the same time. The primary air enters into the surrounding area of the grinding ring uniformly through the nozzle ring, and the mixture of coal powder thrown out tangentially from the grinding ring is dried and transported to the separator on the upper part of the coal mill for separation. The coarse powder is separated and returned to the grinding ring for regrinding [31].

Coal mill is the most important large equipment in coal to oil to powder system. Its operation safety and economy are very important to the normal operation of coal to oil. Specifically, the temperatures of different components in coal mill are important indicators for operation state. Low temperatures will impact the efficiency and the performance of production to some extent, while high temperatures may bring safety risks. Thus, in this case, the temperatures of two components, the motor bearing and the rotary separator bearing, are selected as the quality variables. Our dataset is collected from a real coal mill in power plant, and the detailed information about this dataset is shown in Table I below.



**Fig. 7.** The structure of a medium speed roller type coal mill [31].

TABLE I

The description about coal mill dataset

|  |  |
| --- | --- |
| Sample size | 12000 |
| Basic sampling rate | 1 min |
| Process variables  (34 variables) | * **1 min sampling rate:** coal feed, power, ambient temperature (3 variables) * **2 mins sampling rate:** electric current, rotation rate, pressure, etc. (10 variables) * **10 mins sampling rate:** temperatures of other components (21 variables) |
| Quality variables  (2 variables) | * **20 mins sampling rate:** motor bearing temperature (QV1) and rotary separator bearing temperature (QV2) |

From Table I, the 34 process variables can be divided into 3 groups according to their sampling rates, i.e. 1 min, 2 mins and 10 mins. The two quality variables are sampled at the slowest 20 mins sampling rate, and referred to as QV1 and QV2 for simplicity.

## B. Experiment Setup

According to the time order, the first 8000 samples in the dataset are taken as the train set, while the rest 4000 samples will be taken as the test set. Because the sampling rate of quality variables are 20 times than the basic sampling rate, there will be only 400 and 200 labeled samples in the train set and the test set respectively. The experiment is repeatedly performed 3 times, and the final results are averaged. To evaluate the performance of soft sensor, the coefficient of determination (R2) and Root Mean Squared Error (RMSE), which have been widely used in soft sensor performance evaluation [32,33], are chosen. The computation formulas for R2 and RMSE are given in formula (9) and formula (10) as below.

 (9)

 (10)

where  and  are the ground truth and the prediction of sample .  and  are the size of test set and the average of the ground truth in the test set. The closer R2 is to 1 and the smaller RMSE is, the better the soft sensor performs.

To optimize the hyper-parameters in our proposed model, the latter 2000 samples in the train set are used as validation set, so the grid search technique can be applied to search for the hyper-parameters with best performance on validation set. The candidate ranges of hyper-parameters in MCW-RNN are listed in Table II below. Among them, some hyper-parameters need to be optimized, while some hyper-parameters are fixed based on prior experience.

TABLE II

The candidate ranges of hyper-parameters in MCW-RNN

|  |  |  |
| --- | --- | --- |
| **Hyper-parameter** | | **Candidate range** |
| Optimized | Coefficient for dimension expansion | [15, 30, 60] |
| Dimension of hidden layer in MLP | [64, 128, 256] |
| Weight decay/10-3 | [5, 10, 50, 100] |
| Batch size | [64, 128] |
| Fixed | Learning rate | 0.001 |
| Epochs | 400 |
| Step size for scheduler | 50 |
| Gamma for scheduler | 0.5 |
| Sequence length | 20 |

Besides our proposed MCW-RNN method, six contrast models are designed for comparison. The detailed description about the design of contrast models is given in Table III, where these models belongs to two types, i.e. up-sampling and down-sampling. Because the contrast models are also neural network methods, they use the same optimization strategy for hyper-parameters as shown in Table II. Moreover, all the models, including our proposed MCW-RNN, are trained by Adam optimizer [34], and the step scheduler is used for learning rate adjustment during training.

TABLE III

The description about the design of contrast models

|  |  |  |
| --- | --- | --- |
| **Model type** | **Model name** | **Description** |
| Up-sampling | U-MLP | * Construct MLP after cubic spline interpolation for process variables and quality variables |
| U-RNN | * Construct RNN after cubic spline interpolation for process variables and quality variables |
| Down-sampling | D-MLP | * Construct MLP after down-sampling to the slowest sampling rate |
| D-RNN | * Construct RNN after down-sampling to the slowest sampling rate |
| L-MLP | * Construct MLP after using data lifting technique |
| Z-RNN | * Construct RNN after imputation for process variables with zeros |

## C. Results and Discussions

The performance of our proposed MCW-RNN and contrast models on QV1 and QV2 are respectively shown in Table IV and Table V. “Mean” and “Std” are respectively the mean and standard deviation of the results across 3 times experiments. The best results are shown in the boldface.

TABLE IV

The performance of different models on QV1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **R2 (%)** | | **RMSE (℃)** | |
| **Mean** | **Std** | **Mean** | **Std** |
| U-MLP | 60.02 | 8.55 | 1.022 | 0.110 |
| U-RNN | 60.46 | 13.93 | **1.008** | 0.187 |
| D-MLP | 54.92 | **0.91** | 1.089 | **0.011** |
| D-RNN | 23.53 | 11.07 | 1.275 | 0.091 |
| L-MLP | 38.78 | 4.07 | 1.269 | 0.042 |
| Z-RNN | 55.32 | 3.20 | 1.084 | 0.039 |
| MCW-RNN | **61.35** | 2.01 | **1.008** | 0.026 |

TABLE V

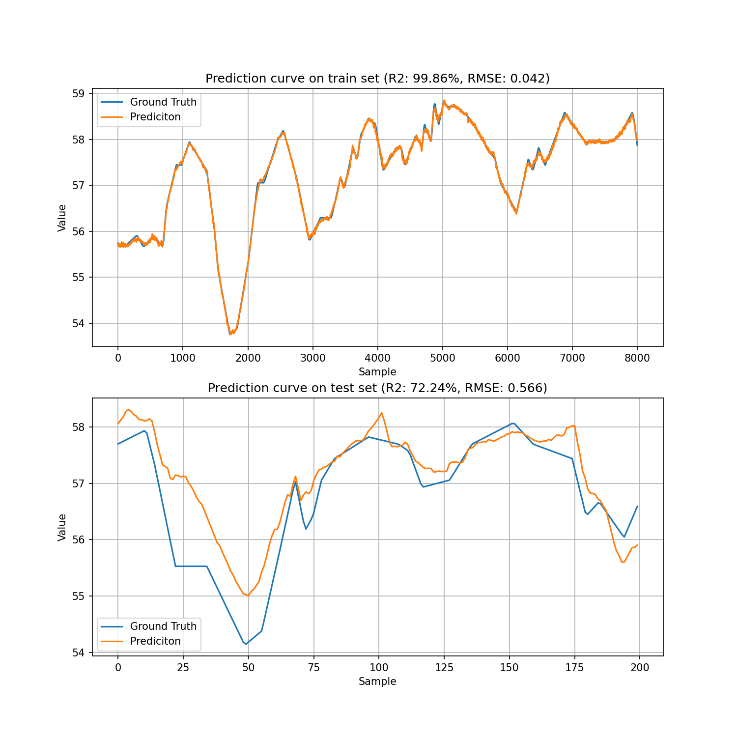
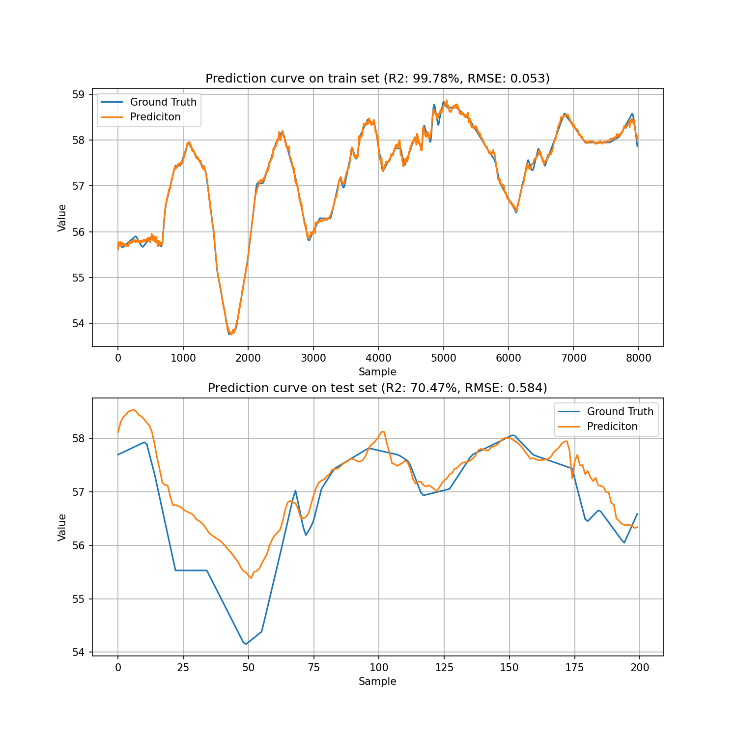
The performance of different models on QV2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **R2 (%)** | | **RMSE (℃)** | |
| **Mean** | **Std** | **Mean** | **Std** |
| U-MLP | 75.09 | 6.66 | 0.533 | 0.075 |
| U-RNN | 78.83 | 7.37 | 0.490 | 0.084 |
| D-MLP | 86.91 | 0.51 | 0.389 | **0.008** |
| D-RNN | 88.82 | 5.09 | 0.359 | 0.080 |
| L-MLP | 84.91 | 0.54 | 0.417 | **0.008** |
| Z-RNN | 86.41 | 1.68 | 0.396 | 0.025 |
| MCW-RNN | **90.59** | **0.49** | **0.329** | 0.009 |

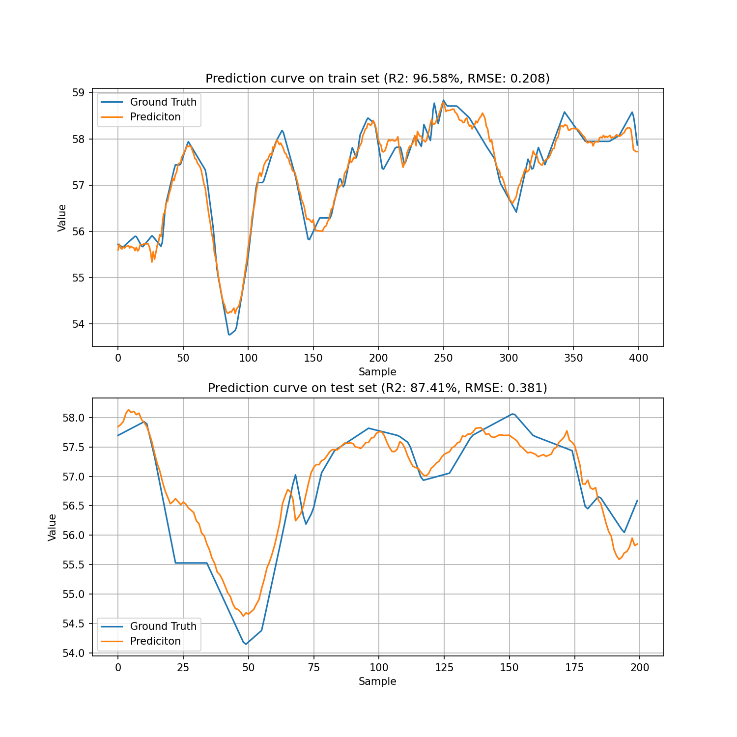
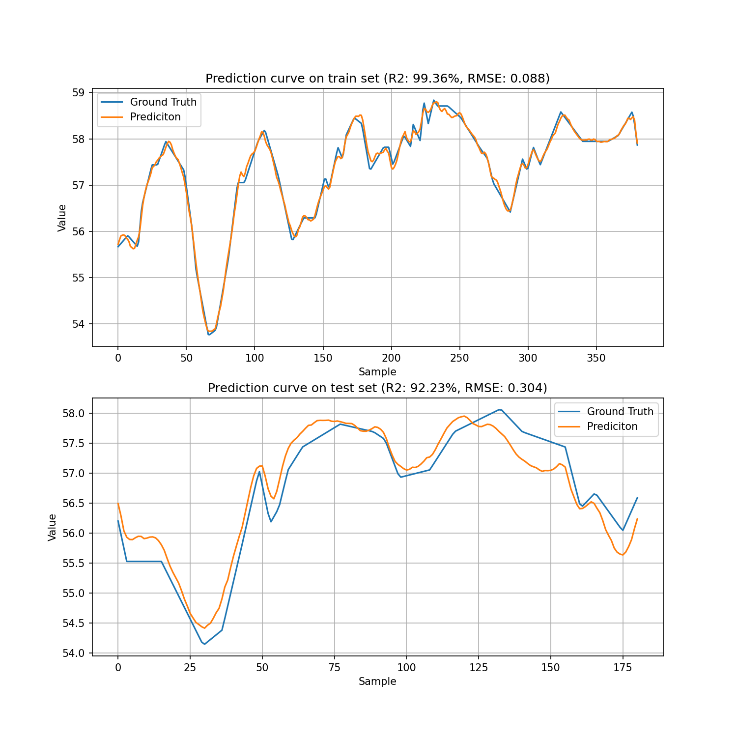
From the mean results shown above, it is obvious that our proposed MCW-RNN outperforms other models on QV1 and QV2 except the same performance as U-RNN on the RMSE index of QV1. Meanwhile, the performance of MCW-RNN is relatively steady, which is reflected from the low standard deviation of the results. As for the up-sampling methods, U-MLP and U-RNN perform similarly that QV1 can be accurately predicted while the performance on QV2 is the worst among all these models. On the whole, it can be easily found that QV2 is much easier to be predicted than QV1. Thus, the up-sampling operation can provide more information to help improve the performance on QV1. While for QV2, the original process variables already possess enough useful information for prediction, so the inaccurate up-sampling operation may introduce interference, which degrades the performance on the contrary. Because the temporal relationship can be learnt by recurrent structure, U-RNN performs slightly better than U-MLP. Among the down-sampling methods, D-MLP and Z-RNN show similar performance. The missing values in Z-RNN are filled with zeros, so process variables are not equally utilized that process variables with faster sampling rate contributes more. D-MLP considers the process variables equally, but it losses the incomplete sample and the temporal information within faster sampling rates. Due to the drawbacks above, D-MLP and Z-RNN are both inferior to MCW-RNN. Although L-MLP takes fully advantage of the data with data lifting technique, the process variables are not equally treated and it neglects the temporal relationship, which results in the poor performance. D-RNN has recurrent structure, but it only captures the temporal relationship at the slowest sampling rate. The temporal relationships within other sampling rates are also helpful for soft sensor, so the performance of D-RNN on QV1 is the worst among these models. Based on the above analysis, our proposed MCW-RNN makes full use of the incomplete samples and is able to capture temporal relationship within different sampling rates. Furthermore, the grouping for hidden neurons and the clockwork updating strategy elaborately memorize and learn the information in process variables with different sampling rates, which equally exploits each process variable.

To further analyze the advantage of MCW-RNN, the prediction curves of different models in once experiment on QV2 are compared in Fig.8. From Fig.8, the prediction curves of different models are roughly the same except the part with green shadow. The prediction curves of U-MLP and U-RNN diverge the most from the ground truth, which is also reflected in the performance index. Compared with U-MLP and U-RNN, L-MLP reduces the divergences quite a lot, and D-MLP and Z-RNN obtains more accurate predictions further. Moreover, D-RNN and MCW-RNN perform the best, and the prediction curve of MCW-RNN nearly perfectly matches the ground truth. Hence, it is demonstrated that our proposed MCW-RNN shows superior performance in some local parts, which are hard to be predicted by other models.

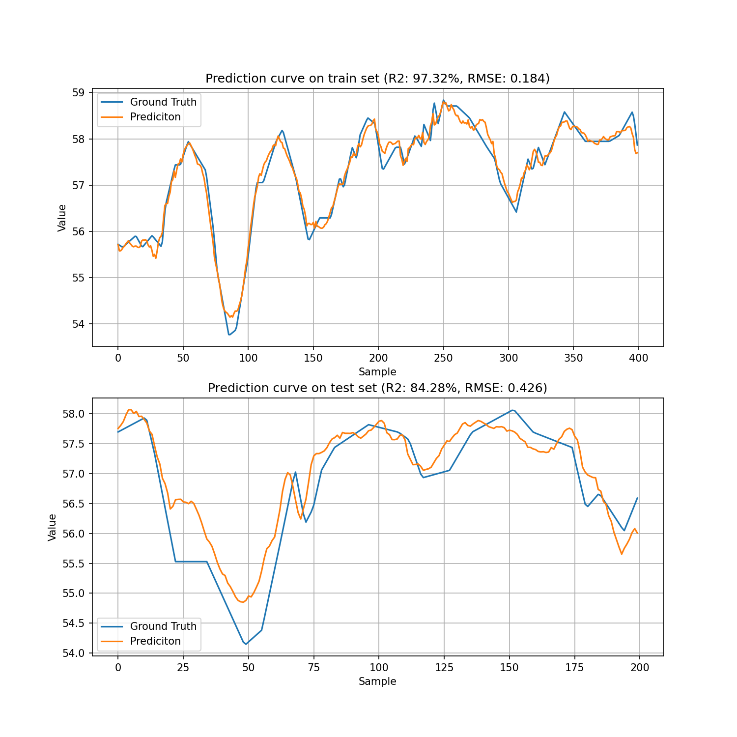
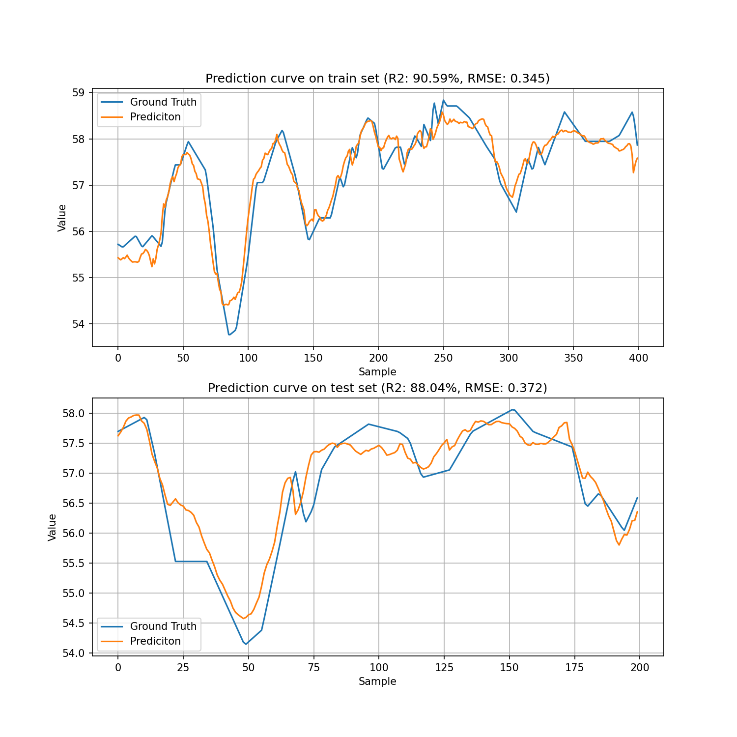
The core idea of MCW-RNN is grouping and updating the hidden neurons according to the constitution of the sampled process variables at each time step. Therefore, it is interesting to analyze how the soft sensor performance evolves along the time step. Because the coal mill dataset is simulated as multirate form, we can evaluate the predictions at each time step with the ground truth. The evolution of MCW-RNN performance along time step on QV2 is illustrated in Fig.9. The sequence length is fixed at 20, so the performance is evaluated at 20 time steps.

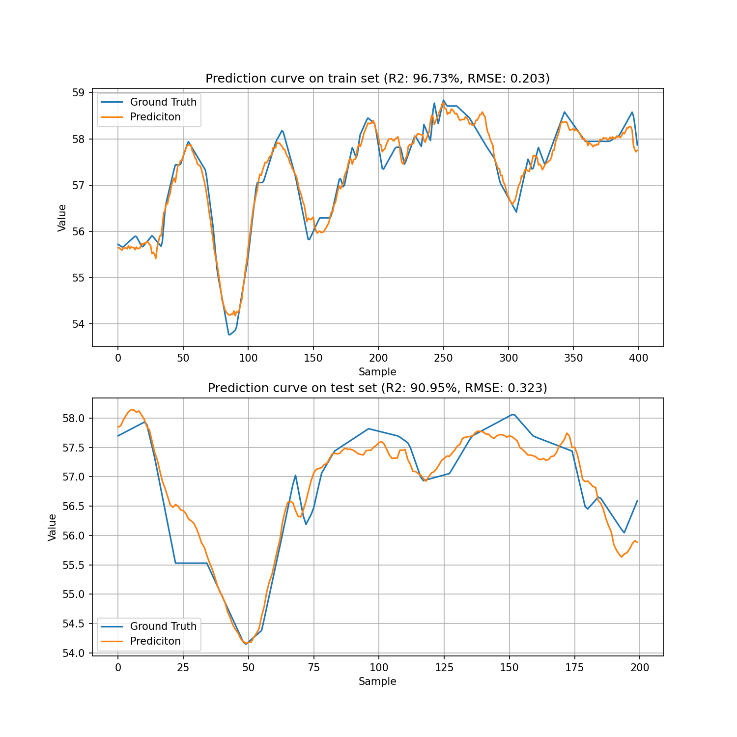
(a) U-MLP (R2(%)=72.24, RMSE(℃)=0.566) (b) U-RNN (R2(%)=70.47, RMSE(℃)=0.584)

(c) D-MLP (R2(%)=87.41, RMSE(℃)=0.381) (d) D-RNN (R2(%)=92.23, RMSE(℃)=0.304)

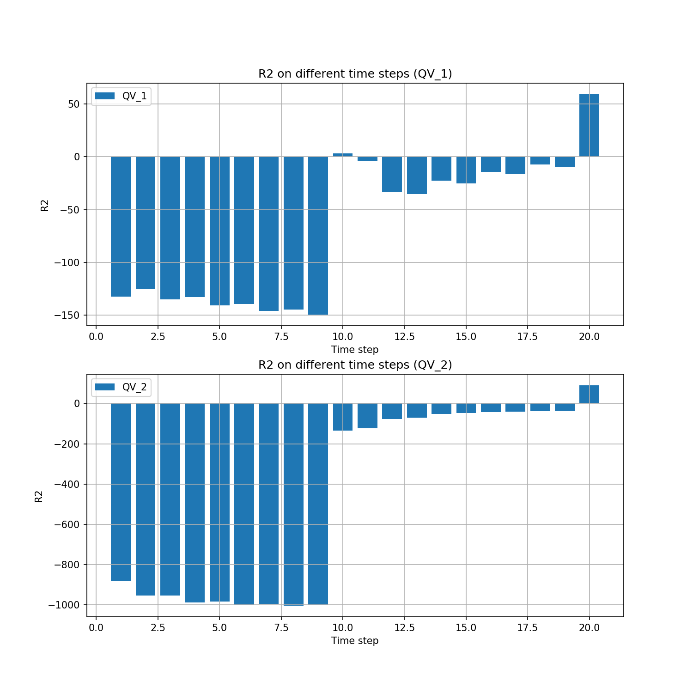
 

(e) L-MLP (R2(%)=84.28, RMSE(℃)=0.426) (f) Z-RNN (R2(%)=88.04, RMSE(℃)=0.372)

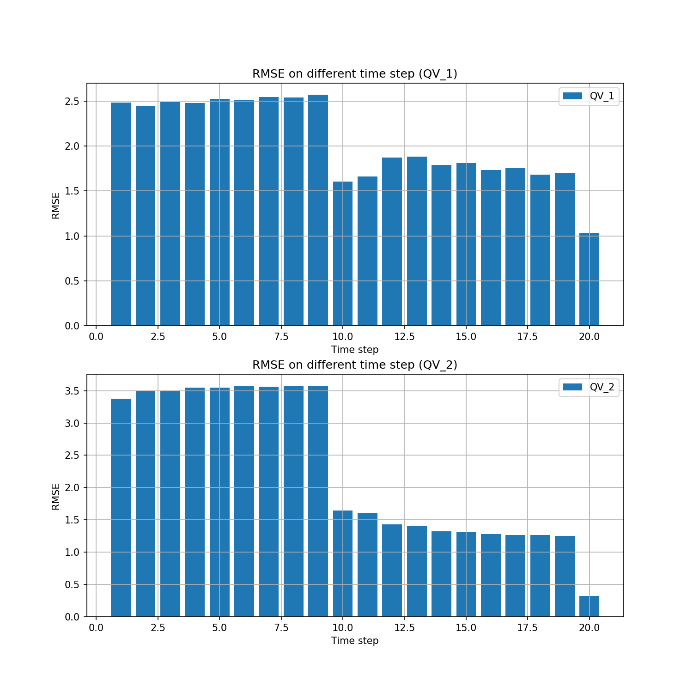


(g) MCW-RNN (R2(%)=90.95, RMSE(℃)=0.323)

**Fig. 8.** The prediction curves of different models in once experiment on QV2.



(a) R2(%)=90.95

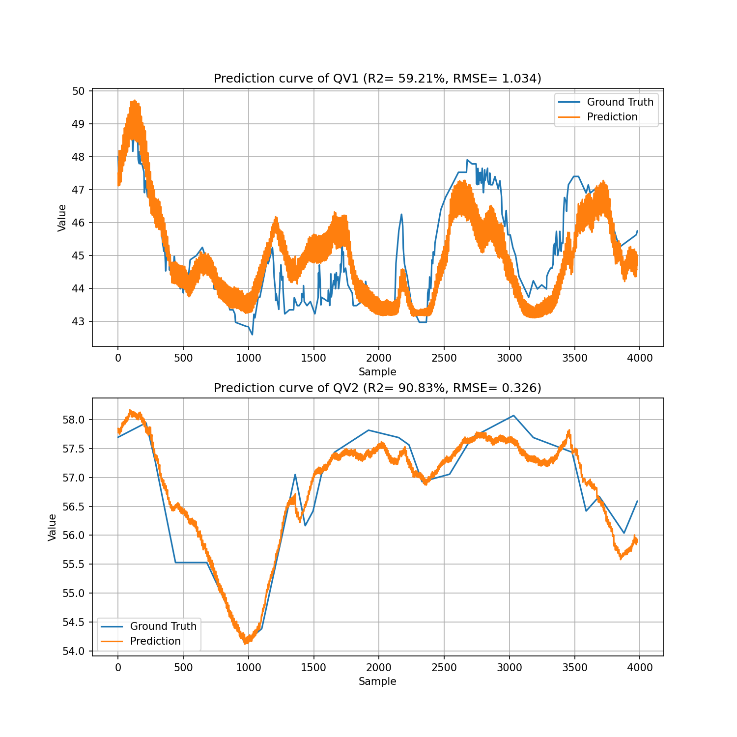


(b) RMSE(℃)=0.323

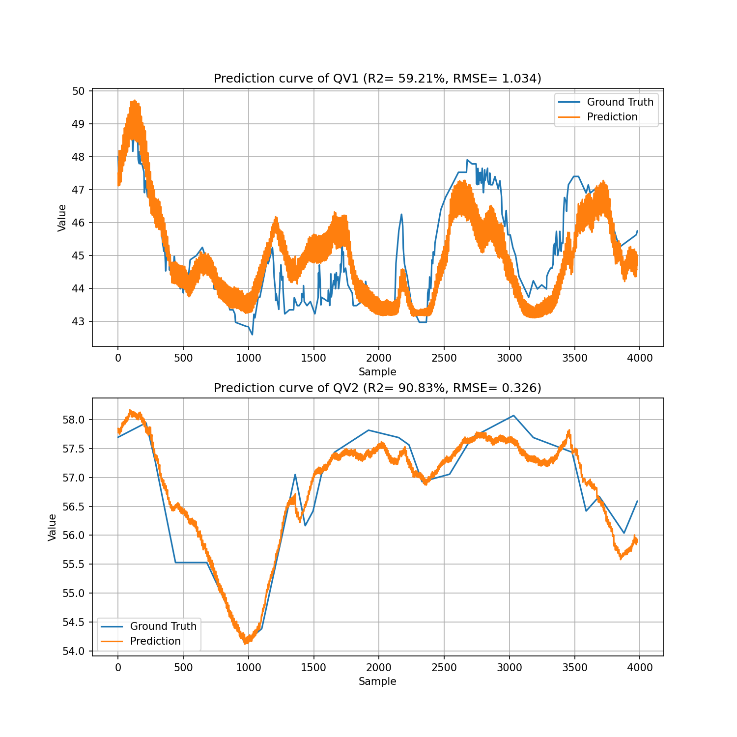
**Fig. 9.** The evolution of MCW-RNN performance along time step on QV2.

As shown in Fig.9, it can be seen that the performance at the first 9 time steps is so poor that the R2 index is lower than 0. Because the temperatures of other components, which are quite relevant to QV2, are sampled per 10 mins, they have not been sampled at the first 9 time steps, which explains the poor performance. Then, at the 10th time step, the process variables with sampling rate of 1 min, 2 mins and 10 mins are all sampled. Thus, the performance shows significant improvement. During the 11th to 19th time steps, the performance is gradually promoted with the enrichment of process variables. Finally, the performance is obviously elevated at the last time step. From the performance evolution along time step, it is shown that the hidden neurons gradually learns the relationship between process variables and quality variables. Also, it is demonstrated that the temperatures of other components are pretty important for the soft sensing of QV2.

As aforementioned, we just care about the prediction performance at the slowest sampling rate. But in practice, the measurement at the slowest sampling rate may not be sufficient for real-time monitoring and control. Thus, it is necessary to prove that whether our proposed MCW-RNN can give reliable and accurate predictions at the fastest sampling rate, i.e. 1 min in this experiment. The prediction curves of MCW-RNN on QV1 and QV2 at the fastest sampling rate are shown in Fig.10.



(a) QV1 (R2(%)=59.21, RMSE(℃)=1.034)



(b) QV2 (R2(%)=90.83, RMSE(℃)=0.326)

**Fig. 10.** The prediction curves of MCW-RNN on QV1 and QV2 at the fastest sampling rate.

In Fig.10, MCW-RNN still performs accurately and steadily, especially on QV2. The results on perform index R2 and RMSE are close to the results in Table IV and Table V, which proves the effectiveness of MCW-RNN at the fastest sampling rate.

From all the results and discussions above, we demonstrate the superiority and effectiveness of the proposed MCW-RNN for multirate soft sensor. With the designed grouping pattern and clockwork updating strategy, MCW-RNN properly adapts to the multirate data structure and utilizes the process variables equally. Besides, the temporal relationships within different sampling rates can be explored, and reliable predictions can be given at any sampling point. As a whole, MCW-RNN implements a simple but promising method for multirate soft sensor modeling.

# V. Conclusions

In this paper, a novel multirate industrial soft sensor, named MCW-RNN, is proposed. Through the hidden neurons grouping and clockwork updating strategy, MCW-RNN effectively balances the utilization for process variables with different sampling rates, and explores the temporal relationships within different sampling rates. Experiments based on real coal mill case demonstrate the excellent performance at the slowest sampling rate and the ability to give reliable predictions at any sampling point. Furthermore, how to introduce the clockwork mechanism for more multirate tasks is a challenging future direction.

# References

1. Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A review on soft sensors for monitoring, control, and optimization of industrial processes," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12868-12881, 2020.
2. Y. Zhao, A. Fatehi, and B. Huang, "Robust estimation of ARX models with time varying time delays using variational Bayesian approach," *IEEE transactions on cybernetics*, vol. 48, no. 2, pp. 532-542, 2017.
3. Z. Chai, C. Zhao, B. Huang, and H. Chen, "A deep probabilistic transfer learning framework for soft sensor modeling with missing data," *IEEE Transactions on Neural Networks Learning Systems*, 2021.
4. Y. Qin, C. Zhao, and B. Huang, "A new soft-sensor algorithm with concurrent consideration of slowness and quality interpretation for dynamic chemical process," *Chemical Engineering Science*, vol. 199, pp. 28-39, 2019.
5. Z. Li, Y.-S. Lee, J. Chen, and Y. Qian, "Developing variable moving window PLS models: Using case of NOx emission prediction of coal-fired power plants," *Fuel*, vol. 296, p. 120441, 2021.
6. C. Zhao, F. Wang, Z. Mao, N. Lu, and M. Jia, "Quality prediction based on phase‐specific average trajectory for batch processes," *AIChE Journal*, vol. 54, no. 3, pp. 693-705, 2008.
7. J. F. Tuttle, R. Vesel, S. Alagarsamy, L. D. Blackburn, and K. Powell, "Sustainable NOx emission reduction at a coal-fired power station through the use of online neural network modeling and particle swarm optimization," *Control Engineering Practice*, vol. 93, p. 104167, 2019.
8. P. Zhou, D. Guo, H. Wang, and T. Chai, "Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking," *IEEE transactions on neural networks learning systems*, vol. 29, no. 9, pp. 4007-4021, 2017.
9. S. Chang, C. Zhao, and K. Li, "Consistent-Contrastive Network with Temporality-Awareness for Robust-to-Anomaly Industrial Soft Sensor," *IEEE Transactions on Instrumentation Measurement*, 2021.
10. H. B. Su, L. Fan, and J. R. Schlup, "Monitoring the process of curing of epoxy/graphite fiber composites with a recurrent neural network as a soft sensor," *Engineering Applications of Artificial Intelligence*, vol. 11, no. 2, pp. 293-306, 1998.
11. X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE transactions on industrial informatics*, vol. 16, no. 5, pp. 3168-3176, 2019.
12. Z. Yong, H. Fang, Y. Zheng, and X. Li, "Torus-event-based fault diagnosis for stochastic multirate time-varying systems with constrained fault," *IEEE Transactions on cybernetics*, vol. 50, no. 6, pp. 2803-2813, 2019.
13. B. Lin, B. Recke, T. M. Schmidt, J. K. Knudsen, and S. B. Jørgensen, "Data-driven soft sensor design with multiple-rate sampled data: A comparative study," *Industrial engineering chemistry research*, vol. 48, no. 11, pp. 5379-5387, 2009.
14. Z. Chai, C. Zhao, and Y. Sun, "A Sequentially-Adaptive Deep Variational Model for Multirate Process Anomaly Detection," in *2021 3rd International Conference on Industrial Artificial Intelligence (IAI)*, 2021, pp. 1-6: IEEE.
15. M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B*, vol. 61, no. 3, pp. 611-622, 1999.
16. Y. Wu, X. Luo, and Z. Yuan, "Soft sensor modeling with dynamic interpolation neutral network for multirate system," *Chemical Industry Engineering Progress*, vol. 28, no. 8, pp. 1323-1327, 2009.
17. Y. Wu and X. Luo, "A novel calibration approach of soft sensor based on multirate data fusion technology," *Journal of Process Control*, vol. 20, no. 10, pp. 1252-1260, 2010.
18. L. Li and Y. Dai, "Dynamic Soft Sensor Development for Time-Varying and Multirate Data Processes Based on Discount and Weighted ARMA Models," *Symmetry*, vol. 11, no. 11, p. 1414, 2019.
19. Z. Chai, C. Zhao, and B. Huang, "Variational progressive-transfer network for soft sensing of multirate industrial processes," *IEEE Transactions on Cybernetics*, 2021.
20. J. Koutnik, K. Greff, F. Gomez, and J. Schmidhuber, "A clockwork rnn," in *International Conference on Machine Learning*, 2014, pp. 1863-1871: PMLR.
21. J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179-211, 1990.
22. K. Cho et al., "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724-1734.
23. C. Olah. (2015). Understanding lstm networks. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
24. S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness Knowledge-Based Systems*, vol. 6, no. 02, pp. 107-116, 1998.
25. S. Kanai, Y. Fujiwara, and S. Iwamura, "Preventing gradient explosions in gated recurrent units," *Advances in neural information processing systems*, vol. 30, pp. 435-444, 2017.
26. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
27. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS 2014 Workshop on Deep Learning*, December 2014, 2014.
28. W. Lu, Z. Yi, W. Liu, Y. Gu, Y. Rui, and B. Ran, "Efficient deep learning based method for multi-lane speed forecasting: a case study in Beijing," *IET Intelligent Transport Systems*, vol. 14, no. 14, pp. 2073-2082, 2021.
29. S. Achanta, T. Godambe, and S. V. Gangashetty, "An investigation of recurrent neural network architectures for statistical parametric speech synthesis," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
30. S. Ruder. (2016). An overview of gradient descent optimization algorithms. Available: <https://arxiv.org/abs/1609.04747>.
31. L. Shanghai Yingyong Machinery Co. (2019). ZGM medium speed coal mill. Available: <http://www.shyychina.com/zgm-medium-speed-coal-mill-15662910362010273.html>.
32. Z. Yao and C. Zhao, "FIGAN: A Missing Industrial Data Imputation Method Customized for Soft Sensor Application," *IEEE Transactions on Automation Science Engineering*, 2021.
33. L. Feng, C. Zhao, Y. Li, M. Zhou, H. Qiao, and C. Fu, "Multichannel diffusion graph convolutional network for the prediction of endpoint composition in the converter steelmaking process," *IEEE Transactions on Instrumentation Measurement*, vol. 70, pp. 1-13, 2020.
34. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference for Learning Representations*, 2015.

**Shuchao Chang** received the B.Eng. degree from Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing the M.Eng. degree with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. His current research interests include industrial soft sensors and machine learning.

**Chunhui Zhao** (SM’15) received the Ph.D. degree from Northeastern University, China, in 2009. From 2009 to 2012, she was a Postdoctoral Fellow with the Hong Kong University of Science and Technology and the University of California, Santa Barbara, Los Angeles, CA, USA.

Since January 2012, she has been a Professor with the College of Control Science and Engineering, Zhejiang University, Hangzhou, China. Her research interests include statistical machine learning and data mining for industrial application. She has authored or coauthored more than 140 papers in peer-reviewed international journals. She has served Senior Editor of Journal of Process Control, AEs of two International Journals, including Control Engineering Practice and Neurocomputing.

**Yingnan Wang** received D.Eng. degree from North China Electric Power University, Beijing, China, in 2021. She is currently a postdoctoral fellow at Zhejiang University, Hangzhou, China. Her current research interests include industrial soft sensors and machine learning.

1. This work is supported by xxx. (The corresponding author is Chunhui Zhao)

   Shuchao Chang, Chunhui Zhao and Yingnan Wang are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China (e-mail: chhzhao@zju.edu.cn). [↑](#footnote-ref-1)