

Article

Temperature Prediction Model for a Regenerative Aluminum Smelting Furnace by a Just-in-Time Learning-Based Triple-Weighted Regularized Extreme Learning Machine

Xingyu Chen, Jiayang Dai *  and Yasong Luo

Guangxi Key Laboratory of Intelligent Control and Maintenance of Power Equipment, School of Electrical Engineering, Guangxi University, Nanning 530004, China

* Correspondence: daijiayang@gxu.edu.cn; Tel.: +86-185-7439-5495

Abstract: In a regenerative aluminum smelting furnace, real-time liquid aluminum temperature measurements are essential for process control. However, it is often very expensive to achieve accurate temperature measurements. To address this issue, a just-in-time learning-based triple-weighted regularized extreme learning machine (JITL-TWRELM) soft sensor modeling method is proposed for liquid aluminum temperature prediction. In this method, a weighted JITL method (WJITL) is adopted for updating the online local models to deal with the process time-varying problem. Moreover, a regularized extreme learning machine model considering both the sample similarities and the variable correlations was established as the local modeling method. The effectiveness of the proposed method is demonstrated in an industrial aluminum smelting process. The results show that the proposed method can meet the requirements of prediction accuracy of the regenerative aluminum smelting furnace.

Keywords: temperature prediction; weighted regularized extreme learning machine; just-in-time learning; sample similarities; variable correlations



Citation: Chen, X.; Dai, J.; Luo, Y. Temperature Prediction Model for a Regenerative Aluminum Smelting Furnace by a Just-in-Time Learning-Based Triple-Weighted Regularized Extreme Learning Machine. *Processes* **2022**, *10*, 1972. <https://doi.org/10.3390/pr10101972>

Academic Editors: Guoqing Zhang, Zejia Zhao and Wai Sze YIP

Received: 1 September 2022

Accepted: 23 September 2022

Published: 30 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Aluminum can be made into alloys with various metals; it is widely used in automotive, aviation, and military industries due to its good ductility, plasticity, recyclability, and oxidation resistance. A regenerative aluminum smelting furnace is important for the aluminum smelting process, in which the real-time measurement and control of liquid aluminum temperatures influence the quality of the aluminum. However, on industrial sites, there are many influencing factors, such as the aging of temperature-measuring thermocouples and fluctuations in the operating voltage, which bring difficulties to the real-time measurements of the aluminum liquid temperature. Hence, it is essential to develop a modeling method to predict the liquid aluminum temperature for quality improvement of the aluminum. The aluminum smelting process is a typical complex industrial furnace production process. In recent decades, many studies on industrial furnaces have been performed (regarding ‘mechanism modeling’) [1–3]. Although the physical meaning of ‘mechanism modeling’ is clear, there are some problems, such as complicated calculations for industrial furnace systems. At the same time, mechanism models may not be reliable enough since they usually make simplified assumptions. The furnace temperature, airflow rate, etc., fluctuate greatly in different working states due to the intermittent working characteristics of the regenerative aluminum smelting furnace. The real-time update of the model for the regenerative aluminum smelting furnace is also a problem that needs to be considered.

To overcome the shortcomings of mechanism modeling, a soft-sensor that makes full use of the industrial data is proposed [4]. There are many researchers working on the data-driven modeling of industrial furnaces and similar processes, such as partial

least squares (PLS) [5], the kernel principal component regression (KPCR) [6], and kernel partial least squares (KPLS) [7], which have been successfully applied with good results. However, these methods are generally considered to be global modeling (and trained offline). Moreover, after these models are put into application, they will face problems, such as difficulties in model updating. Consequently, to deal with the adaptive update problem of the model, the moving window technique [8,9], recursive models [10,11], and the just-in-time learning (JITL) strategy [12,13] are usually used as online adaptive update strategies. The JITL strategy trains an online local model to predict the query samples by selecting similar samples from historical samples, so it is more suitable for processes such as industrial furnaces with state mutations. For example, Chen et al. [14] proposed a least squares support vector machine temperature prediction model based on JITL to deal with large temperature change lags in roller kilns. Dai et al. [15] combined the moving window technique and the JITL strategy as an update strategy to select similar samples in both time and space dimensions, and they verified the effectiveness of the proposed method on an industrial kiln. In [16], a locally weighted partial least squares regression (LWPLS) model was proposed by JITL-based local modeling. In LWPLS, the samples most similar to the query sample are assigned different weights and selected for local modeling. The current model will be discarded when the next query sample is available. Then, a new local PLS model will be established for the model's online update. However, LWPLS only considers the sample similarities, not the variable correlations. The data of the aluminum smelting process often present high-dimensional characteristics and each input variable has a different degree of influence on the liquid aluminum temperature. Hence, except for the sample similarities, it is necessary to consider the variable correlations [17–19]. Furthermore, the accuracy of the JITL strategy depends on the quality of the selected samples. However, the traditional similarity measurement criteria, such as Euclidean distance and Mahalanobis distance, only consider the input information without considering the output information, and often cannot obtain accurate similar samples. Thus, investigating new similarity measurement criteria is important for the JITL strategy.

In recent years, artificial intelligence algorithms, such as long short-term memory networks (LSTM) [20–22] and extreme learning machine(s) (ELM) [23–26] have also been used in soft sensor modeling. The basic assumption for LSTM is that process data are sampled at even and unified frequencies; it is very difficult to meet these conditions for 'process data measurements' in industrial processes, especially for quality variables. Hence, LSTM is unsuitable for some processes with irregular sampling frequencies. ELM is a single hidden layer neural network with a low algorithm complexity, which does not need backpropagation to solve iteratively, and has been used in the temperature prediction of regenerative aluminum smelting processes. Huang et al. [27] proposed an extreme learning machine furnace temperature prediction model based on the kernel principal component analysis and showed that ELM has a better effect than the traditional BP neural network. Liu et al. [28] proposed an ELM model optimized by the restricted Boltzmann machine (RBM) to solve the random initialization of the input weights and biases in the ELM. Moreover, ELM has a fast learning speed and is suitable as an online prediction model. For example, Li et al. [29] built a local online ELM model in combination with a JITL strategy, allowing the online prediction of polyethylene terephthalate (PET) viscosity without relying on time-consuming laboratory analysis procedures. However, this ELM-based online prediction model neither considers sample similarities nor variable correlations, which is unreasonable in local modeling. Moreover, the original ELM runs the risk of model overfitting. Hence, a regularized extreme learning machine (RELM) [30] was proposed to solve the model's overfitting problem.

Although some research studies have been carried out on ELM, there are few discussions about sample similarities and variable correlations in RELM, especially in temperature prediction. Based on the above discussions, a soft sensor modeling method of the JITL-based triple-weighted regularized extreme learning machine (JITL-TWRELM) was proposed to solve the above problems. Compared with the traditional data-driven

modeling method described above, the method proposed in this paper not only allows real-time updating of the model but also obtains more accurate local modeling samples due to the use of the WJITL strategy, which uses correlation information between the input and output variables in the sample selection stage. Meanwhile, in the local modeling stage, the proposed method overcomes the shortcomings of the traditional local modeling method, which only considers the sample similarities and analyzes the variable correlations, highlighting the influences of different variables on the output. The remainder of this article is structured as follows. Firstly, the regenerative aluminum smelting furnace is briefly introduced. Secondly, the regularized extreme learning machine (RELM), sample weighted regularized extreme learning machine (SWRELM), and variable weighted regularized extreme learning machine (VWRELM) are introduced, respectively. Then, the JITL-based triple-weighted regularized extreme learning machine (JITL-TWRELM) is described. Next, the flexibility and effectiveness of the proposed method are validated in the industrial aluminum smelting processing. Finally, we present the conclusions.

2. Related Methods

Since ELM runs the risk of model overfitting, the regularization method is used to solve the overfitting problem. Considering sample similarities and variable correlations, the sample weighted regularized extreme learning machine (SWRELM) and the variable weighted extreme learning machine (VWRELM) are introduced, respectively. Three related methods are discussed next. To better understand the derivation of the relevant equations, the definition of symbols in this paper is shown in Table 1.

Table 1. Definition of symbols in this paper.

Symbols	Definition
x_n, t_n	the n th historical input and output variable vectors
β_i	the output weight of the i th hidden layer unit
$\beta, \beta^S, \beta^V, \beta^t$	the output weight vectors in RELM, SWRELM, VWRELM, JITL-TWRELM
T	the output vector of RELM
ω_i, b_i	the input weight and bias connecting input layer and i th hidden layer unit
\hat{t}_j, \hat{t}_q	the output corresponding to x_j , the output of the query sample in JITL-TWRELM
N	the number of training samples
C	the regularization coefficient
ξ	the training error vector
H, H^V, H^t	the hidden layer output matrices in RELM, VWRELM, JITL-TWRELM
$\Omega_{sn}, \Omega_s, \Omega_s^t$	the sample weight of the n th sample, the sample weighted matrix, the sample weighted matrix in JITL-TWRELM
λ	the Lagrange multiplier vector
ρ	the Pearson correlation coefficient
$E(x), E(t)$	the expectation of the single input variable and output variable
v_i	the contribution of each variable
V	the variable contribution matrix
x_n^v, x_n^w	the variable weighted input sample, the variable weighted local modeling sample
$d_{on}, d_{on}^w, d_{on}^{tw}$	the original Euclidean distance and weighted Euclidean distance, the weighted Euclidean distance in JITL-TWRELM
x_q, x_q^w	the query sample, the variable weighted query sample in JITL-TWRELM
$\Omega_v, \Omega_v^l, \Omega_v^g$	the correlation coefficient matrix, the local correlation coefficient matrix, the global correlation coefficient matrix
φ	the adjusted parameter
X, X^w	the local modeling sample matrix, the variable weighted local modeling sample matrix in JITL-TWRELM

2.1. RELM

As shown in Figure 1, the structure of ELM consists of three parts, which are the input layer, hidden layer, and output layer [31]. The core idea of ELM is to randomly select the input weights and hidden layer biases of the network. The output weights between the hidden layer and output layer are obtained by minimizing the loss function and solving the Moore–Penrose generalized inverse operation. Owing to the particularity of the single hidden-layer structure, ELM has a faster learning speed, minimal human

interference, and it is easier to implement than traditional networks. However, the original ELM model only considers the empirical risk minimization (ERM) principle, which tends to result in an overfitting model. To overcome this deficiency, a regularized extreme learning machine (RELM) was proposed based on empirical risk minimization and structural risk minimization (SRM) principles and has proven to be a better generalization performance than ELM.

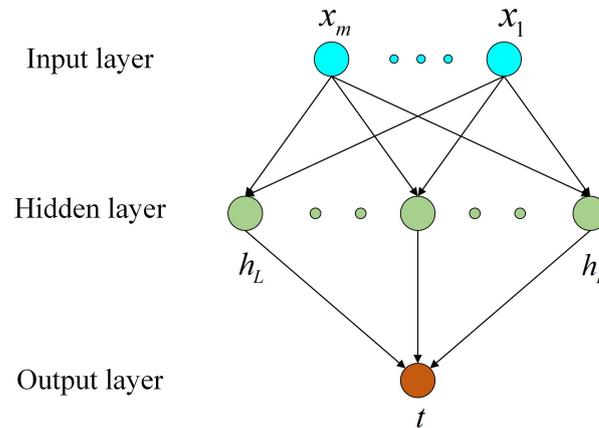


Figure 1. The structure of ELM.

It is assumed that the n th historical input variable vector and the output variable are denoted as $x_n = [x_{n1}, x_{n2}, \dots, x_{nm}]$ and t_n , respectively, where m is the number of input variables. (x_n, t_n) is the n th historical sample composed of x_n and t_n . The output function of the RELM with L hidden layer neurons can be represented as

$$\sum_{i=1}^L \beta_i g(\omega_i x_j^T + b_i) = t_j, j = 1, \dots, N \quad (1)$$

where β_i is the output weight of the i th hidden layer unit, $\omega_i = [\omega_{i1}, \dots, \omega_{im}]$, and b_i are the input weight, bias connecting input layer, and i th hidden layer unit, respectively. $x_j = [x_{j1}, x_{j2}, \dots, x_{jm}]$ is the input variable vector, t_j denotes the output corresponding to x_j , N is the number of training samples. $g(\cdot)$ is the activation function. Usually, $g(\cdot)$ is set as the sigmoid function. We re-write Equation (1) in matrix form

$$H\beta = T \quad (2)$$

where

$$H = [h(x_1^T)^T, \dots, h(x_N^T)^T]^T = \begin{pmatrix} g(\omega_1 x_1^T + b_1) & \dots & g(\omega_L x_1^T + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N^T + b_1) & \dots & g(\omega_L x_N^T + b_L) \end{pmatrix} \quad (3)$$

$$\beta = [\beta_1, \dots, \beta_L]^T \quad (4)$$

$$T = [t_1, \dots, t_N]^T \quad (5)$$

Due to ω_i and b_i being randomly given, to obtain the output weight vector β , the optimization equation can be represented as

$$\begin{aligned} \min & \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \|\xi\|^2, \\ \text{s.t.} & h(x_j^T)\beta = t_j + \xi_j, j = 1, \dots, N \end{aligned} \quad (6)$$

where C represents the regularization coefficient, which can adjust the empirical risk and structural risk. $\xi = [\xi_1, \dots, \xi_N]^T$ is the training error vector. By constructing the Lagrange function, the solution of Equation (6) is

$$\beta = \begin{cases} (H^T H + \frac{I_L}{C})^{-1} H^T T, L < N \\ H^T (H^T H + \frac{I_N}{C})^{-1} T, L > N \end{cases} \quad (7)$$

where $I_L \in R^{L \times L}$, $I_N \in R^{N \times N}$.

2.2. SWRELM

Not all samples have the same contribution to the output; moreover, the original RELM considers all samples equally important and does not consider the differences between different samples. Thus, to obtain a more realistic result, the sample weighted matrix $\Omega_s = \text{diag}(\Omega_{s1}, \dots, \Omega_{sN})$ is added to Equation (6), which is expressed as

$$\begin{aligned} \min & \frac{1}{2} \|\beta^S\|^2 + \frac{C}{2} \|\Omega_s \xi\|^2, \\ \text{s.t.} & h(x_j^T) \beta^S = t_j + \xi_j, j = 1, \dots, N \end{aligned} \quad (8)$$

The Lagrange function can be represented as follows:

$$\begin{aligned} L(\beta^S, \Omega_s, \lambda) &= \frac{1}{2} \|\beta^S\|^2 + \frac{C}{2} \|\Omega_s \xi\|^2 - \sum_{j=1}^N \lambda (\sum_{i=1}^L h(x_j^T) \beta^S - t_j - \xi_j) \\ &= \frac{1}{2} \|\beta^S\|^2 + \frac{C}{2} \|\Omega_s \xi\|^2 - \lambda (H \beta^S - T - \xi) \end{aligned} \quad (9)$$

where $\lambda = [\lambda_1, \dots, \lambda_N]$ denotes the Lagrange multiplier vector. According to the KKT condition, taking the derivative of Equation (9) and setting the derivative to zero, we have

$$\frac{\partial L}{\partial \beta^S} = 0 \rightarrow (\beta^S)^T = \lambda H \quad (10)$$

$$\frac{\partial L}{\partial \xi} = 0 \rightarrow C \xi^T \Omega_s^2 + \lambda = 0 \quad (11)$$

$$\frac{\partial L}{\partial \lambda} = 0 \rightarrow H \beta^S = T + \xi \quad (12)$$

With Equations (11) and (12), the Lagrange multiplier vector λ can be expressed as

$$\lambda = -C(H \beta^S - T)^T \Omega_s^2 \quad (13)$$

Similarly, with Equations (10) and (13), the expression of the output weight vector of the sample weighted regularized extreme learning machine (SWRELM) is

$$\beta^S = (H^T \Omega_s^2 H + \frac{I_L}{C})^{-1} H^T \Omega_s^2 T, L < N \quad (14)$$

Equation (14) is suitable when the number of modeling samples is greater than the number of hidden neurons. Moreover, in this case, β^S has a faster calculation speed [32].

2.3. VWRELM

The original RELM treats all input variables with equal importance, while not all input variables have the same effect on the output variable, some input variables are more strongly correlated with the output variable than others. Thus, to reflect the differences of input variables and obtain better quality-related features, a variable contribution method

based on the Pearson correlation coefficient was adopted. On this basis, the variable weighted extreme learning machine (VWRELM) was proposed. The Pearson correlation coefficient is defined as

$$\rho = \frac{E(xt) - E(x)E(t)}{\sqrt{E(x^2) - E^2(x)}\sqrt{E(t^2) - E^2(t)}} \quad (15)$$

where $E(x)$ and $E(t)$ are the expectations of the single input variable and output variable, respectively. ρ represents the degree of correlation between the two variables; two highly correlated variables will also have a larger ρ . As a result, the variable contribution can be defined by ρ . For a training sample $(x_n, t_n), n = 1, \dots, k$, where each input sample x_n has m dimensions, the contribution of each variable can be defined as

$$v_i = \frac{|\rho_i|}{\sum_{j=1}^m |\rho_j|}, i = 1, \dots, m \quad (16)$$

where ρ_i represents the Pearson correlation coefficient between the i th input variable and the output variable. The variable contribution matrix can be written as

$$V = \text{diag}(v_1, \dots, v_m) \quad (17)$$

Hence, taking the variable contribution as the variable weights, and applying the variable weights to the input sample x_n , the weighted input sample can be expressed as

$$x_n^v = x_n V = x_n \text{diag}(v_1, \dots, v_m) = (x_{n1}v_1, \dots, x_{nm}v_m) \quad (18)$$

where x_n^v represents the input sample weighted by variable weights. It can be seen from Equation (18) that each dimension of the input sample is given a different weight, reflecting the differences between variables. By variable weighting, Equation (3) can be rewritten as

$$\begin{aligned} H^V &= [h((x_1^v)^T)^T, \dots, h((x_N^v)^T)^T]^T \\ &= \begin{pmatrix} g(\omega_1((x_1^v)^T)^T + b_1) & \dots & g(\omega_L((x_1^v)^T)^T + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1((x_N^v)^T)^T + b_1) & \dots & g(\omega_L((x_N^v)^T)^T + b_L) \end{pmatrix} \end{aligned} \quad (19)$$

when $L < N$, the output weight vector is

$$\beta^V = ((H^V)^T H^V + \frac{I_L}{C})^{-1} (H^V)^T T, L < N \quad (20)$$

3. The Proposed JITL-TWRELM Model

In the previous analysis, the RELM, SWRELM, and VWRELM models have been established. However, in a multi-data, multivariate prediction model, the different samples and variables to the predicted outputs are different, especially in the aluminum smelting process. Table 2 shows the shortcomings of the three methods. Both sample similarities and variable correlations should be taken into account in RELM. Hence, to obtain a better model, combined with the weighted JITL strategy (WJITL), a JITL-based triple-weighted regularized extreme learning machine is proposed.

Table 2. Shortcomings of the three methods.

Method	Shortcomings
RELM	Neither sample similarities nor variable correlations are considered, and the model cannot be updated in real-time.
SWRELM	Only the sample similarities are considered, no variable correlations are considered, and the model cannot be updated in real-time.
VWRELM	Only the variable correlations are considered, no sample similarities are considered, and the model cannot be updated in real-time.

3.1. Weighted Similarity Measurement Criterion

The original Euclidean distance is usually used as a similarity measurement criterion, expressed as

$$d_{on} = \sqrt{(x_q - x_n)(x_q - x_n)^T} \quad (21)$$

where $x_q \in R^{1 \times m}$ is the current query sample, $x_n \in R^{1 \times m}$ is the n th historical sample, and d_{on} indicates the Euclidean distance between the current query sample and the n th historical sample. The more similar the historical sample is to the query sample, the smaller the distance d_{on} . However, Equation (21) only uses the input information of the historical sample and query sample, while the information of the output is not taken into consideration. Moreover, the calculation of the Euclidean distance can be regarded as the accumulation of each dimension of the sample. It is easy to see that the importance of each dimension may be different, with some dimensions contributing more to distance than others. Hence, inspired by Equation (15), the connections between the input variables and output variables are established through the correlation analysis. We define a weighted Euclidean distance as a weighted similarity measure criterion, expressed as

$$d_{on}^w = \sqrt{(x_q - x_n)\Omega_v(x_q - x_n)^T} \quad (22)$$

where $\Omega_v = \text{diag}(\rho_1, \dots, \rho_m)$. Then, the sample weight is expressed as

$$\Omega_{sn} = \exp\left(\frac{d_{on}^w}{\varphi^2}\right) \quad (23)$$

where φ is the adjust parameter, which can adjust the change rate of weight value with the sample distance. For a better expression, the JITL strategy that applied this weighted similarity measurement criterion is called WJITL.

3.2. JITL-TWRELM

A JITL-based triple-weighted regularized extreme learning machine (JITL-TWRELM) soft sensor method, combined with the WJITL strategy, was established to simultaneously incorporate sample weights and variable weights. The detailed derivation steps are as follows.

$N(N < H)$ samples $(x_n, t_n), n = 1, \dots, H$ from historical samples were selected for each query sample to local modeling. First, Pearson correlation coefficients between input variables and output variables of all historical samples were calculated to obtain the correlation coefficient matrix

$$\Omega_v^g = \text{diag}(\rho_1^g, \dots, \rho_m^g) \quad (24)$$

To distinguish it from the subsequent derivation, we call Ω_v^g the global correlation coefficient matrix, where $\rho_i^g, i = 1, \dots, m$ is the global correlation coefficient. As a result, the

weighted Euclidean distance between the query samples and the historical samples can be obtained by Equation (25).

$$d_{on}^{tw} = \sqrt{(x_q - x_n)\Omega_v^g(x_q - x_n)^T}, n = 1, \dots, H \quad (25)$$

We sort $d_{on}^{tw}, n = 1, \dots, H$ from small to large, and the first N samples are selected as modeling samples. The sample weighted matrix is obtained as

$$\Omega_s^t = \text{diag}(\exp(\frac{d_{o1}^{tw}}{\varphi^2}), \dots, \exp(\frac{d_{oN}^{tw}}{\varphi^2})) = \text{diag}(\Omega_{s1}^t, \dots, \Omega_{sN}^t) \quad (26)$$

Then, the Pearson correlation coefficient of N local modeling samples is calculated, and the local correlation coefficient matrix is obtained as

$$\Omega_v^l = \text{diag}(\rho_1^l, \dots, \rho_m^l) \quad (27)$$

where Ω_v^l is used as the local variable weighted matrix for local modeling samples and the query sample

$$X^w = X\Omega_v^l = \{x_n^w\}, n = 1, \dots, N \quad (28)$$

$$x_q^w = x_q\Omega_v^l \quad (29)$$

where $X \in R^{N \times m}$ consists of local modeling samples, X^w and x_q^w are the variable weighted local modeling sample and variable weighted query sample, respectively. Thus, the new local modeling dataset $(x_n^w, t_n), n = 1, \dots, N$ is used to build the local model. The optimization equation for the output weight vector is established as Equation (30)

$$\begin{aligned} \min & \frac{1}{2} \|\beta^t\|^2 + \frac{C}{2} \|\Omega_s^t \xi\|^2, \\ \text{s.t.} & h((x_j^w)^T) \beta^t = t_j + \xi_j, j = 1, \dots, N \end{aligned} \quad (30)$$

The output matrix of the hidden layer is

$$H^t = \begin{pmatrix} g(\omega_1(x_1^w)^T + b_1) & \dots & g(\omega_L(x_1^w)^T + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1(x_N^w)^T + b_1) & \dots & g(\omega_L(x_N^w)^T + b_L) \end{pmatrix} \quad (31)$$

According to Equations (9)–(14), the output weight vector of JITL-TWRELM is

$$\beta^t = ((H^t)^T (\Omega_s^t)^2 H^t + \frac{I_L}{C})^{-1} (H^t)^T (\Omega_s^t)^2 T, L < N \quad (32)$$

Finally, the prediction output of the query sample is

$$\hat{t}_q^t = \sum_{i=1}^L \beta_i^t g(\omega_i(x_q^w)^T + b_i) \quad (33)$$

4. Industrial Case

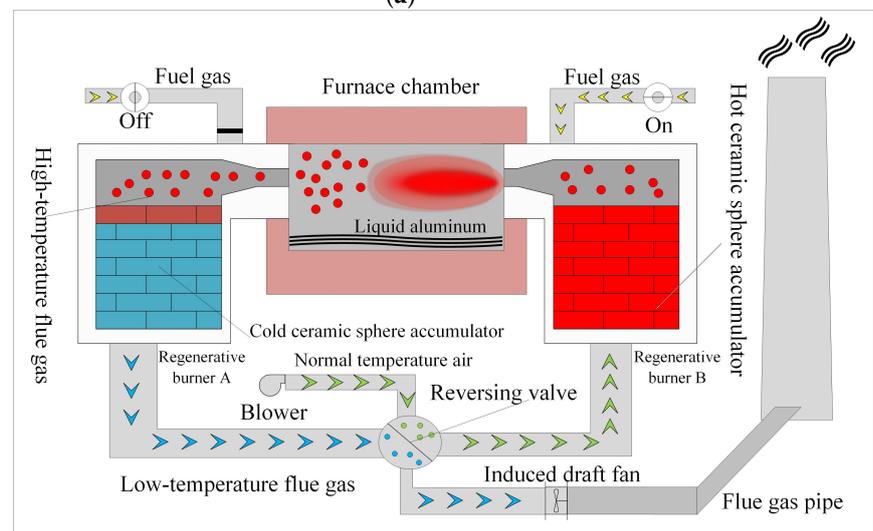
4.1. Process Description of the Regenerative Aluminum Smelting Furnace

An industrial regenerative aluminum smelting furnace and its internal structure are shown in Figure 2a and Figure 2b, respectively. The regenerative aluminum smelting furnace consists of a furnace chamber, regenerative burner (including burner and ceramic sphere accumulator), reversing valve, flue gas pipe, etc. The regenerative burners are arranged in pairs, and the two opposite burners are a group (A and B). Normal temperature air from the blower enters burner B through the reversing valve and is heated as it flows through the hot ceramic sphere accumulator. Then, the normal temperature air is heated to

a temperature close to the furnace chamber (generally 80% to 90% of the furnace chamber temperature). The heated high-temperature air enters the furnace chamber and then rolls up the flue gas around the furnace to form a thin oxygen-poor high-temperature airflow with an oxygen content lower than 21%. Then, the mixture of the oxygen-poor high-temperature air and the injected flue gas is ignited to smelt the aluminum material. At the same time, the high-temperature flue gas passes through burner A, the heat is stored in the cold ceramic sphere accumulator, and then the flue gas is discharged at a temperature lower than 150 °C through the flue gas pipe. When the stored heat reaches saturation, the reversing valve is reversed, and the regenerative burner A and B change their combustion and heat storage working states, and so on, resulting in energy savings (and reducing emissions).



(a)



(b)

Figure 2. (a) An industrial regenerative aluminum smelting furnace; (b) the internal structure of the regenerative aluminum smelting furnace.

4.2. Model Establishment

To construct the model for the prediction of the liquid aluminum temperature, 12 secondary variables were chosen as the input variables, which are shown in Table 3. These input variables were measured by the sensor. The measurement ranges and errors of the sensors are shown in Table 4. The sampling interval of each sampling point was five

minutes. There were 4400 data samples collected for modeling, of which, 4000 samples were used as historical data for training, and 400 samples for model testing. To better test the effectiveness of the proposed method, two groups of data (D1 and D2) from different periods were used as the testing dataset, with 200 samples in each group.

Table 3. Input variables for the model in the aluminum smelting process.

Input	Variable
1	Material temperature
2	Furnace pressure
3	12 # combustion airflow
4	12 # combustion air pressure difference
5	34 # combustion airflow
6	34 # combustion air temperature
7	34 # combustion air pressure difference
8	34 # gas air-fuel ratio
9	B1 # exhaust gas temperature
10	B2 # exhaust gas temperature
11	B3 # exhaust gas temperature
12	B4 # combustion air temperature

Table 4. Sensor measurement range and error.

Sensor Type	Measurement Range	Measurement Error
Pressure meter	0–15,000 Pa	1%
Flow meter	0–15 m ³ /h	1.5%
Thermocouple	0–1300 °C	1%

The flowchart of JITL-TWRELM model is shown in Figure 3. To validate the performance of JITL-TWRELM, the six methods listed below were employed for comparison.

- Method 1: JITL-RELM (it applies the original JITL strategy and original RELM).
- Method 2: JITL-SWRELM (it applies the original JITL strategy and sample weights on RELM).
- Method 3: WJITL-RELM (it applies the WJITL strategy and original RELM).
- Method 4: JITL-VWRELM (it applies the original JITL strategy and local variable weights on RELM).
- Method 5: JITL-DWRELM (it applies the WJITL strategy and sample weights on RELM).
- Method 6: JITL-TWRELM (it applies the WJITL strategy, sample weights, and local variable weights on RELM).

The detailed step-by-step procedure of the proposed method is as follows.

Step 1: Prepare the input and output variables of the historical samples and perform the standardization.

Step 2: Determine the number N of training samples selected from the total historical samples, the parameter φ for the sample weight calculation, the hidden neuron number L , and the regularization coefficient C of the regularized extreme learning machine.

Step 3: Analyze the global correlation between the input variables and output variables of all historical samples. The global correlation coefficient matrix Ω_v^g is calculated for the sample similarity measurement.

Step 4: Calculate the weighted Euclidean distances between the current query samples and the training samples; N samples closest to the current query sample are selected as local modeling samples.

Step 5: Analyze the local correlation between input variables and output variables of the local modeling samples. The local correlation coefficient matrix Ω_v^l is determined.

Step 6: The JITL-TWRELM model is established, and the output of the current query sample is predicted.

Step 7: Before the next query sample arrives, the previous model is discarded and a new model is constructed based on the next query sample, enabling real-time updating of the model.

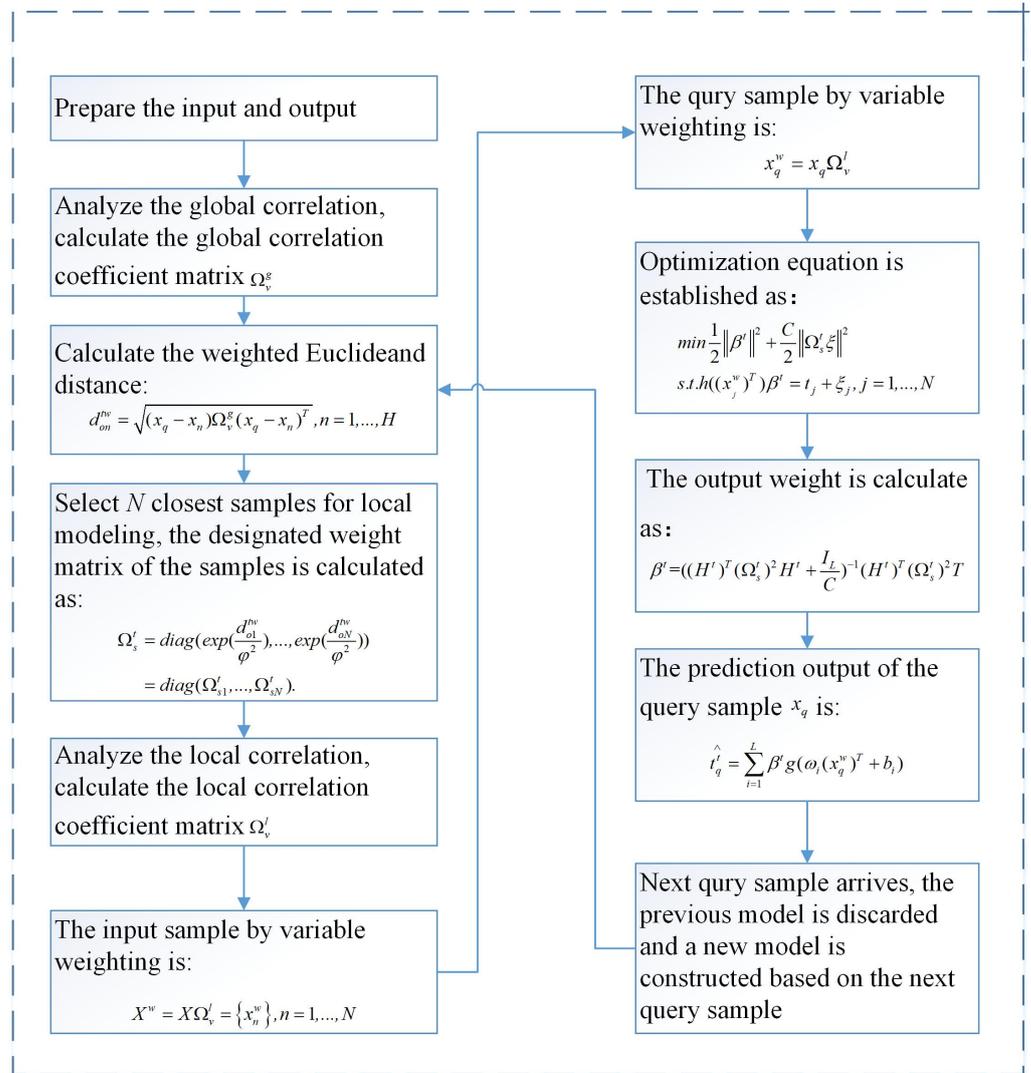


Figure 3. The flowchart of JITL-TWRELM model.

To evaluate the performance of the proposed method, four indices, including mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) are used in the performance evaluation, which are as follows:

$$MAE = \frac{1}{N_T} \sum_{i=1}^{N_T} |y_i - \hat{y}_i| \tag{34}$$

$$RMSE = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2} \tag{35}$$

$$MAPE = \frac{1}{N_T} \sum_{i=1}^{N_T} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{36}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_T} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_T} (y_i - \bar{y})^2} \quad (37)$$

where N_T denotes the number of samples used for testing, y_i and \hat{y}_i denote the values of the actual output variable and predicted output, respectively, \bar{y} denotes the mean value of the actual output variable. It is essential to have small *MAE*, *RMSE*, and *MAPE*, and large R^2 for a prediction model.

Before establishing the JITL-TWRELM model, four parameters need to be determined. By trial and error experiments on dataset D1, N was set as a proper value of 200, which has a good prediction accuracy without increasing the computational burden. Similarly, the parameters φ and L are set to 0.3 and 20, respectively. Table 5 shows the prediction accuracy of the model under the different regularization coefficients. It can be seen that when $C = 150$, the model has a better effect.

Table 5. Comparison of the modeling accuracy with C .

C	<i>MAE</i>	<i>RMSE</i>	<i>MAPE</i>	R^2
140	15.0884	18.6443	0.020354	0.98666
150	14.7273	17.9456	0.019897	0.98764
160	15.3797	19.5019	0.020879	0.98541
170	15.5217	20.4519	0.021023	0.98395
180	15.8944	20.1194	0.021629	0.98447
190	16.0543	19.8318	0.021839	0.98491
200	17.1959	22.2015	0.023265	0.98109

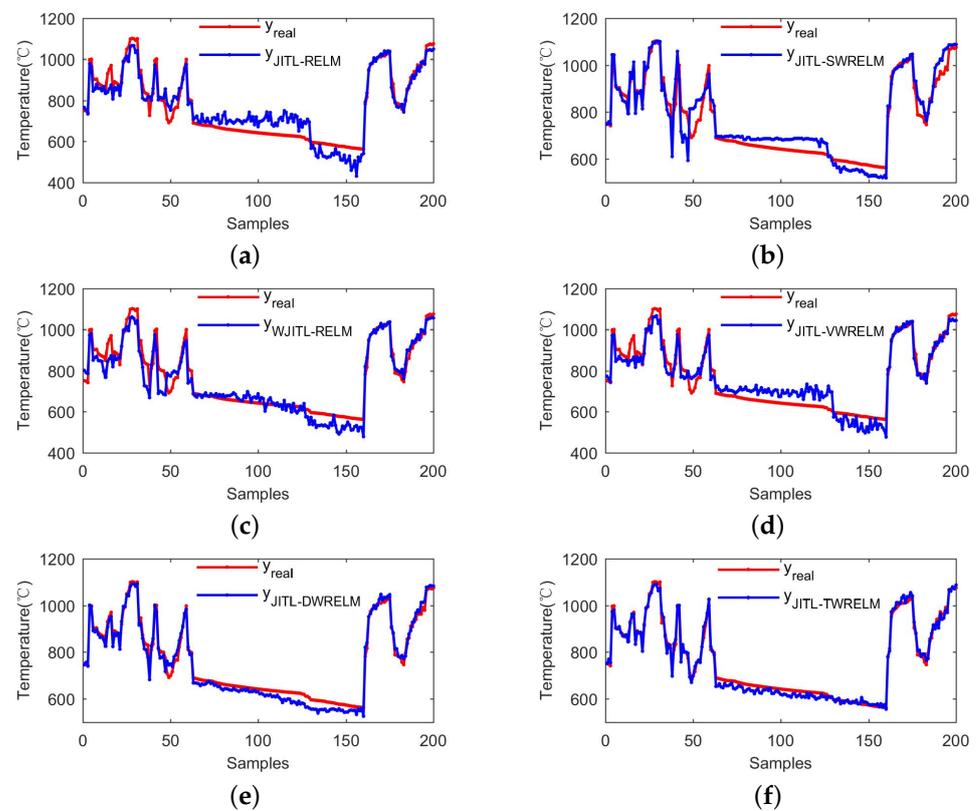
4.3. Results and Discussion

To reduce the effect of randomness on the results, we took the average of ten tests as the final result. The prediction error indices of the six methods on two groups of the testing samples are shown in Table 6. We use testing dataset D1 as an example; in general, the proposed method (method 6) performed better than the other five methods on all four indices. Despite using the JITL strategy, the original method (method 1) had the worst performance on all indices. Methods 2, 3, 4, and 5 also achieved higher prediction accuracies than method 1, as neither the sample weights nor variable weights were used in method 1. Method 2 emphasizes the importance of the samples and introduces sample weights to reflect the effects of different samples on the output. Contrasted with the original JITL strategy, method 3 uses a weighted similarity measurement criterion; samples that are more similar to the query sample were selected to set up the local model, resulting in a more accurate prediction. Different from the previous methods, method 4 considers the local variable weights before establishing the model; the variable weights can be used to improve the influence of output-related variables and reduce that of irrelevant variables in feature extraction [33]. Although methods 2 to 4 have good prediction accuracy improvements, these methods only consider certain types of weighting strategies, such as individual sample weights or variable weights. Hence, method 5 introduces the sample weights and the WJITL strategy, and the R^2 is improved from 0.89453 to 0.97690 compared with method 1. Meanwhile, based on methods 4 and 5—method 6 has the smallest *MAE*, *RMSE*, and *MAPE*, and the highest R^2 among all methods. The R^2 of method 6 is improved from 0.97690 to 0.98764 compared with method 5. Correspondingly, the proposed method 6 has good prediction accuracy on D2, the R^2 reached 0.97427, which is 0.072 higher than method 1.

Table 6. The indices of the six methods of two groups of testing datasets.

Dataset	Method	MAE	RMSE	MAPE	R ²
D1	JITL-RELM	43.0278	52.4279	0.062519	0.89453
	JITL-SWRELM	38.1265	47.9605	0.052589	0.91174
	WJITL-RELM	32.7444	42.7606	0.044443	0.92984
	JITL-VWRELM	38.0149	46.1055	0.054197	0.91843
	JITL-DWRELM	20.7980	24.5347	0.029768	0.97690
	JITL-TWRELM	14.7273	17.9456	0.019897	0.98764
D2	JITL-RELM	26.7981	36.1509	0.035878	0.90223
	JITL-SWRELM	26.3624	33.526	0.03657	0.91174
	WJITL-RELM	27.4121	34.5595	0.044443	0.91065
	JITL-VWRELM	24.3605	31.4511	0.032461	0.92600
	JITL-DWRELM	16.2472	22.2734	0.021632	0.96289
	JITL-TWRELM	14.8733	18.5463	0.019646	0.97427

To more intuitively demonstrate the performances of these six methods, the detailed prediction results for each method on D1 and D2 are shown in Figures 4 and 5, in which (a–f) shows the prediction results of the six methods, respectively. It is easy to see that the prediction of JITL-TWRELM matches well with the curve of the actual measurement of the furnace temperature, while the prediction curve of JITL-RELM cannot track with the real output curve in some samples. In addition, although the other four methods have certain improvements, they still do not achieve the desired effects. In summary, the flexibility and effectiveness of the proposed methods are validated.

**Figure 4.** The detailed prediction results of six methods on D1; (a) method 1; (b) method 2; (c) method 3; (d) method 4; (e) method 5; (f) method 6.

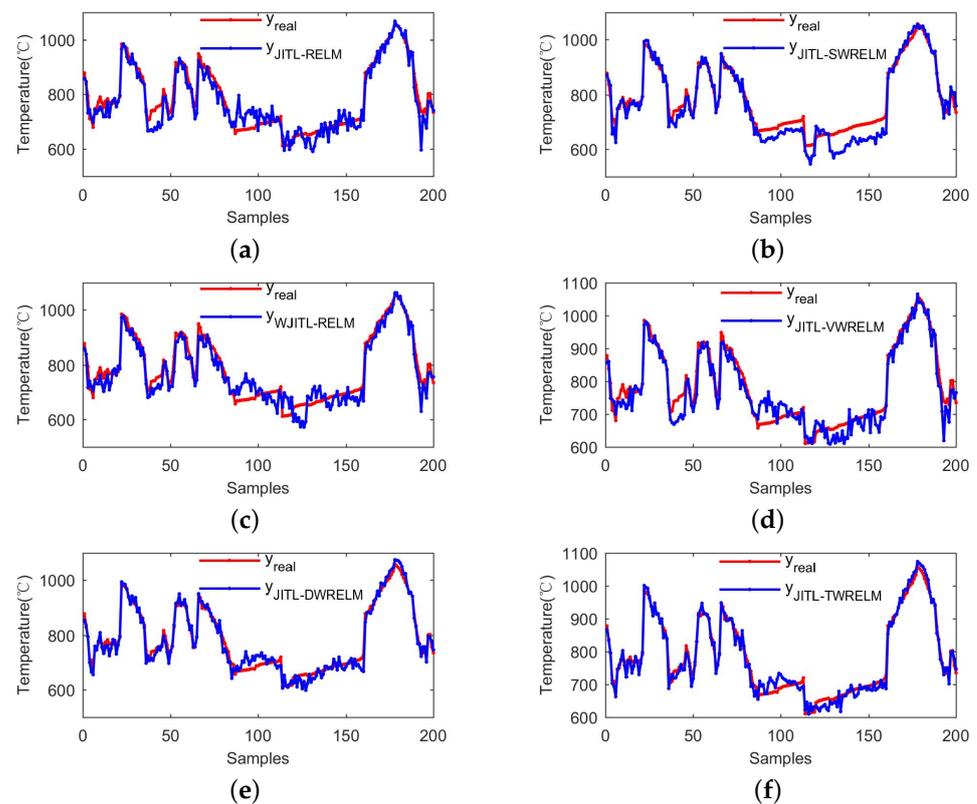


Figure 5. The detailed prediction results of six methods on D2; (a) method 1; (b) method 2; (c) method 3; (d) method 4; (e) method 5; (f) method 6.

The liquid aluminum temperature of the regenerative smelting furnace is generally controlled by feedback. The thermocouple is set in the furnace chamber, and if the temperature is detected to be lower than the set value, the regenerative burner starts to work. In a real industrial site, the temperature measurement performance of the thermocouple used to measure the temperature of aluminum liquids is often affected by voltage fluctuations and the aging of the protective jacket. Old thermocouples need to be replaced frequently, resulting in increased costs. The proposed method in this paper only requires the establishment of a historical database in the industrial site, and whenever a new query sample arrives, the modeling sample is selected from the historical database for modeling, and the prediction results of the aluminum liquid temperature can be obtained. As can be seen from Table 6, the MAEs of the proposed method 6 are 14.7273 and 14.8733 for the two test sets, respectively. Comparing the temperature range and measurement error of the thermocouple in Table 4, the accuracy of the proposed soft measurement model is close to the actual sensor, with a close to 2% error at the maximum temperature measurement range, but the efficiency and costs are more advantageous than the sensor. Therefore, the method proposed in this paper is significant for reducing production costs and improving product quality.

5. Conclusions

This paper mainly deals with the estimation of the liquid aluminum temperature in the regenerative aluminum smelting furnace. A JITL-TWRELM soft sensor modeling method is proposed. In this method, both the sample similarities and the variable correlations are considered in RELM to deal with the differences between samples and variables. Each modeling sample is assigned different weights according to the similarity calculation, and each dimension of the sample is also assigned a corresponding weight according to the correlation analysis, which improves the accuracy of the modeling compared with the original RELM. Furthermore, a weighted similarity measurement criterion is proposed for

JITL to select similar samples for local modeling. Compared with the original JITL strategy, more similar modeling samples are selected for each query sample, enhancing the accuracy and reliability of the local modeling dataset. The flexibility and effectiveness of JITL-TWRELM were validated through the industrial aluminum smelting process. The industrial applications show that the proposed method can effectively deal with the nonlinear and time-varying problems in the regenerative aluminum smelting process and achieve a higher accuracy of temperature prediction compared with the other five methods.

For each query sample, the model needs to be updated once, although some adjacent query samples do not need to update the model so frequently. Selective updating of the model will improve the modeling efficiency. Therefore, developing a selective update strategy will be the focus of future work.

Author Contributions: Data curation, Y.L.; methodology, J.D.; supervision, J.D.; validation, X.C.; writing—original draft, X.C.; writing—review and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there are no conflict of interest regarding the publication of this paper.

References

1. Froehlich, C.; Strommer, S.; Steinboeck, A.; Niederer, M.; Kugi, A. Modeling of the media supply of gas burners of an industrial furnace. *IEEE Trans. Ind. Appl.* **2016**, *52*, 2664–2672. [\[CrossRef\]](#)
2. Strommer, S.; Niederer, M.; Steinboeck, A.; Kugi, A. A mathematical model of a direct-fired continuous strip annealing furnace. *Int. J. Heat Mass Trans.* **2014**, *69*, 375–389. [\[CrossRef\]](#)
3. Qiu, L.; Feng, Y.H.; Chen, Z.G.; Li, Y.L.; Zhang, X.X. Numerical simulation and optimization of the melting process for the regenerative aluminum melting furnace. *Appl. Therm. Eng.* **2018**, *145*, 315–327. [\[CrossRef\]](#)
4. Li, D.Y.; Song, Z.H. A novel incremental gaussian mixture regression and its application for time-varying multimodal process quality prediction. In Proceedings of the 2020 IEEE 9th Data Driven Control and Learning Systems Conference, Liuzhou, China, 19–21 June 2020.
5. Bao, L.; Yuan, X.F.; Ge, Z.Q. Co-training partial least squares model for semi-supervised soft sensor development. *Chemometr. Intell. Lab.* **2015**, *147*, 75–85. [\[CrossRef\]](#)
6. Wang, G.; Luo, H.; Peng, K.X. Quality-related fault detection using linear and nonlinear principal component regression. *J. Franklin. Inst.* **2016**, *353*, 2159–2177. [\[CrossRef\]](#)
7. Liu, H.B.; Yang, C.; Huang, M.Z.; ChangKyoo, Y. Soft sensor modeling of industrial process data using kernel latent variables-based relevance vector machine. *Appl. Soft. Comput.* **2020**, *90*, 106149. [\[CrossRef\]](#)
8. Shi, X.; Huang, G.L.; Hao, X.C.; Yang, Y.; Li, Z. A Synchronous Prediction Model Based on Multi-Channel CNN with Moving Window for Coal and Electricity Consumption in Cement Calcination Process. *Sensors* **2021**, *21*, 4284. [\[CrossRef\]](#)
9. Zhe, L.; Yi-Shan, L.; Chen, J.H.; Qian, Y.W. Developing variable moving window PLS models: Using case of NO_x emission prediction of coal-fired power plants. *Fuel* **2021**, *296*, 120441–120456.
10. Mai, T.; Yu, X.L.; Gao, S.; Frejinger, E. Routing policy choice prediction in a stochastic network: Recursive model and solution algorithm. *Transport. Res. B-Methodol.* **2021**, *151*, 42–58. [\[CrossRef\]](#)
11. Doshi, P.; Gmytrasiewicz, P.; Durfee, E. Recursively modeling other agents for decision making: A research perspective. *Artif. Intell.* **2020**, *279*, 103202–103220. [\[CrossRef\]](#)
12. Qiu, K.P.; Wang, J.L.; Zhou, X.J.; Guo, Y.Q.; Wang, R.T. Soft sensor framework based on semisupervised just-in-time relevance vector regression for multiphase batch processes with unlabeled data. *Ind. Eng. Chem. Res.* **2020**, *59*, 19633–19642. [\[CrossRef\]](#)
13. Wu, Y.J.; Liu, D.J.; Yuan, X.F.; Wang, Y.L. A just-in-time fine-tuning framework for deep learning of SAE in adaptive data-driven modeling of time-varying industrial processes. *IEEE. Sens. J.* **2020**, *21*, 3497–3505. [\[CrossRef\]](#)
14. Chen, N.; Luo, L.H.; Gui, W.H.; Guo, Y.Q. Integrated modeling for roller kiln temperature prediction. In Proceedings of the 2017 Chinese Automation Congress, Shandong, China, 20–22 October 2017.
15. Dai, J.Y.; Chen, N.; Yuan, X.F.; Gui W.H.; Luo, L.H. Temperature prediction for roller kiln based on hybrid first-principle model and data-driven MW-DLWKPCR model. *ISA Trans.* **2020**, *98*, 403–417. [\[CrossRef\]](#)
16. Yao, L.; Ge, Z.Q. Locally weighted prediction methods for latent factor analysis with supervised and semisupervised process data. *IEEE Trans. Autom. Sci. Eng.* **2016**, *14*, 126–138. [\[CrossRef\]](#)
17. Chen, J.M.; Yang, C.H.; Zhou, C.; Li, Y.G.; Zhu, H.Q.; Gui, W.H. Multivariate regression model for industrial process measurement based on double locally weighted partial least squares. *IEEE. Trans. Instrum. Meas.* **2019**, *69*, 3962–3971. [\[CrossRef\]](#)

18. Yuan, X.F.; Huang, B.; Ge, Z.Q.; Song, Z.H. Double locally weighted principal component regression for soft sensor with sample selection under supervised latent structure. *Chemometr. Intell. Lab.* **2018**, *153*, 116–125. [[CrossRef](#)]
19. Lin, W.L.; Hang, H.F.; Zhuang, Y.P.; Zhang, S.L. Variable selection in partial least squares with the weighted variable contribution to the first singular value of the covariance matrix. *Chemometr. Intell. Lab.* **2018**, *183*, 113–121. [[CrossRef](#)]
20. Yuan, X.F.; Li, L.; Shardt, Y.; Wang, Y.L.; Yang, C.H. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Trans. Ind. Electron.* **2020**, *65*, 4404–4414. [[CrossRef](#)]
21. Yuan, X.F.; Li, L.; Wang, Y.L. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 3168–3176. [[CrossRef](#)]
22. He, Z.S.; Chen, Y.H.; Xu, J. A Combined Model Based on the Social Cognitive Optimization Algorithm for Wind Speed Forecasting. *Processes* **2022**, *10*, 689. [[CrossRef](#)]
23. Wang, X.L.; Zhang, H.; Wang, Y.L.; Yang, S.M. ELM-Based AFL–SLFN Modeling and Multiscale Model-Modification Strategy for Online Prediction. *Processes* **2019**, *7*, 893. [[CrossRef](#)]
24. Li, G.Q.; Chen, B.; Qi, X.B.; Zhang, L. Circular convolution parallel extreme learning machine for modeling boiler efficiency for a 300 MW CFBB. *Soft. Comput.* **2019**, *23*, 6567–6577. [[CrossRef](#)]
25. Hu, P.H.; Tang, C.X.; Zhao, L.C.; Liu, S.L.; Dang, X.M. Research on measurement method of spherical joint rotation angle based on ELM artificial neural network and eddy current sensor. *IEEE Sens. J.* **2021**, *21*, 12269–12275. [[CrossRef](#)]
26. Su, X.L.; Zhang, S.; Yin, Y.X.; Xiao, W.D. Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2739–2752. [[CrossRef](#)]
27. Huang, Q.B.; Lei, S.N.; Jiang, C.L.; Xu, C.H. Furnace Temperature Prediction of Aluminum Smelting Furnace Based on KPCA-ELM. In Proceedings of the 2018 Chinese Automation Congress, Xi'an, China, 30 November–2 December 2018.
28. Liu, Q.; Wei, J.; Lei, S.; Huang, Q.B.; Zhang, M.Q.; Zhou, X.B. Temperature prediction modeling and control parameter optimization based on data driven. In Proceedings of the 2020 IEEE Fifth International Conference on Data Science in Cyberspace, Hong Kong, China, 24–27 June 2020.
29. Li, Z.X.; Hao, K.R.; Chen, L.; Ding, Y.S.; Huang, B. Pet viscosity prediction using jit-based extreme learning machine. In Proceedings of the 2018 IFAC Conference, Changchun, China, 20–22 September 2018.
30. Wang, Y.; Li, Y.G.; Wu, B.Y. Improved regularized extreme learning machine short-term wind speed prediction based on gray correlation analysis. *Wind. Eng.* **2021**, *45*, 667–679.
31. Pan, Z.Z.; Meng, Z.; Chen, Z. J; Gao, W.Q.; Shi, Y. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. *Mech. Syst. Signal Process.* **2020**, *144*, 106899–106916. [[CrossRef](#)]
32. Zong, W.W.; Huang, G.B.; Chen, Y.Q. Weighted extreme learning machine for imbalance learning. *Neurocomputing* **2013**, *101*, 229–242. [[CrossRef](#)]
33. Chen, N.; Dai, J.Y.; Yuan, X.F.; Gui, W.H.; Ren, W.T.; Koivo, H. Temperature prediction model for roller kiln by ALD-based double locally weighted kernel principal component regression. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2001–2010. [[CrossRef](#)]