A Novel Superheat Identification of Aluminum Electrolysis with Kernel Semi-supervised Extreme Learning Machine

Ye Jiang\(^1\), Peiqiang Li\(^1\) and Yongxiang Lei\(^2\)*

\(^1\) College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China
\(^2\) School of Automation, Central South University, Changsha, 410083, China
E-mail: leiyongxiang@csu.edu.cn

Abstract. In the aluminium reduction production industry, the superheat temperature (ST) is a vital index, which gives the distribution of the current efficiency. Keeping ST in an approximate range improves the lifespan of the electrolysis bath. However, in the practical production process, ST identification result is commonly evaluated by the experimental workers, the real-time measurement of ST is still a challenge that has not been solved. A novel ST measurement method based on kernel extreme learning machine (K-ELM) is studied in this paper. First, a few of input variables are selected according to expert experience. Then, a new activation function (IG kernel function) and regularization term are constructed in ELM to construct IG-SSELM. Finally, the ST model is built with all the dataset samples and further applied to ST real-time detection. The proposed method is applied to ST identification for the first time in the aluminium industrial process which is superior to the existing ST methods in accuracy and robustness.

1. Introduction

ST is a critical index in the aluminum electrolysis production, maintaining ST in a dynamic interval has a significant effect on the overall production process. A proper ST situation can maintain a good furnace shape, improve the current efficiency, and lengthen aluminum reduction cell lifespan. A lot of ST detection methods have been introduced in the past decades. For example, Chen et al. [1] utilize the frame hole image to extract features to detect ST, however, the number of images is very few, and collecting the image is expensive which cannot realize ST online identification. Zeng et al. [2] analyze the elements that effects the ST and provide a control method. Chen et al. [3] use the fuzzy semantic network to the aluminum electrolysis condition identification, which promotes the research of ST in the industrial aluminum production.

Although some ST method has been applied to ST detection in aluminum electrolysis process. However, the real-time measurement of ST is still a problem that has not yet been solved. Recently, ELM is an efficient and effective algorithm for training the single hidden layers [4-5]. ELM has two main advantages: (1) Random mapping, initialize the input weight between input layers, and output layers. (2) Output weights solving, by solving a matrix that we can get the closed solution of the output weight matrix. ELM has gradually extended and further applied to many other domains such as sparse representation and classification [6], computer vision [7] and clustering learning and image processing, etc. [8]. Semi-supervised learning has been proposed to leverage the information contained in unlabeled patterns. On the other hand, a manifold regularization term was added in the model training for learning the geometry structure of data distribution. Manifold regularization can upgrade the accuracy rate and
generation ability proposed by Belkin [9] which utilized in the ELM training process. To overcome the shortage of judgment made experienced operators and workers, a new soft sensor method based on kernel semi-supervised ELM is proposed to identify ST in this paper. We also propose a novel activation function to replace the Gaussian function to avoid the outlier weight imbalance problem. The main framework of the IG-SSELM is presented as follows. First, Select the appropriate characteristics based on expert experience. Then, utilize all the collected data (labelled and unlabeled data) to train our model with the proposed kernel activation function [10]. Finally, for evaluating the performance of IG-SSELM, which further applied ST identification to industrial aluminum reduction cell. The main advantage of IG-SSELM is described:

- A novel kernel activation function is proposed in the ST identification model.
- All the dataset (labelled and unlabeled data) was used in the training process which improves the accuracy.
- Laplacian regularization is introduced to ELM for obtaining better robustness and generation ability.

The rest of the paper is organized as follows. Section 2 introduces the related works of ELM, SS-ELM, and manifold regularization, which constructs the foundation of the proposed ST model. Section 3 further proposed the main algorithm framework. Industrial experiments are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Works
In this Section, we introduce ELM, Laplacian regularization and SS-ELM framework, which lay the foundation of our later research.

2.1. ELM
ELM is a single-hidden-layer neural network (SLFN) with high training performance [4]. \( \{X, Y\} = \{X, Y\}^N_{i=1} \) is the training data with \( N \) samples. Define the input of the ELM with \( x_i \in \mathbb{R}^{n_x} \), and \( y_j \in \mathbb{R}^{n_y} \) is the output class label, where \( n_i \) and \( n_0 \) are input and output neurons. The detailed structure of ELM is stated in figure 1. A common equation in ELM is:

\[
\sum_{i=1}^{N_h} \gamma i \varsigma (a_i, x_i, b_i) = a_j
\]

(1)

where \( N_h \) represents the hidden nodes and \( a_i \) is the weight matrix. It connects the input layer and the \( i \)th hidden layer. Equation (1) can be also rewritten with a matrix form \( H\gamma = Y \).

\[
H = \begin{bmatrix}
\varsigma (a_1, x_1, b_1) & \cdots & \varsigma (a_{N_h}, x_1, b_{N_h}) \\
\vdots & \ddots & \vdots \\
\varsigma (a_1, x_{N_h}, b_1) & \cdots & \varsigma (a_{N_h}, x_{N_h}, b_{N_h})
\end{bmatrix}, \quad \gamma = \begin{bmatrix}
\gamma_1 \\
\vdots \\
\gamma_{N_h}
\end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix}
y_1 \\
\vdots \\
y_{N_h}
\end{bmatrix}
\]

SLFNs is to find the appropriate \( a, b \) and \( \gamma \) that satisfy the following equation:

\[
\left\| H(a, b)\gamma - Y \right\| = \min_{a,b,\gamma} \| H(a, b)\gamma - Y \|
\]

(2)

Equation (2) can be solved with gradient-based method. ELM learns the hidden weights by a pseudo-inverse matrix, and avoids the BP iterative methods, so the learning efficiency is high.

A continuous probability distribution is used to compute the hidden weights. The hidden output layer of ELM is \( H(x_i) \in \mathbb{R}^{1 \times n_h} \), where \( n_h \) denotes the neurons of the hidden layer. The network then is described by

\[
y_i = H(x_i)\gamma \quad i = 1, 2, \ldots, N
\]

(3)
where $\gamma \in \mathbb{R}^{n_h \times n_0}$ is the output weight matrix. It connects the hidden layer and the output layer.

To obtain the output weights, the objective loss function of ELM is:

$$
\min_{y \in \mathbb{R}^{n_0 \times n}} \Gamma_{ELM} = \frac{1}{2} \|y\|^2 + \frac{C}{2} \|y - H_{\gamma}\|^2
$$

(4)

where $H=[H(x_1)^T, ..., H(x_N)^T]^T \in \mathbb{R}^{N \times n_h}$, $C$ is the relative penalty coefficient, respectively. By setting the gradient of loss function with respect to output weights to 0. As in Ref. [11], the output matrix is calculated with the following equation:

$$
\gamma^* = (J^T J + \frac{I_{n_h}}{C})^{-1} J^T Y
$$

(5)

where $I_{n_h}$ denotes the values of an identity matrix. The ELM structure is described as figure 1.

2.2. SS-ELM

In semi-supervised ELM model, manifold regularization is considered as a regularization term added in ELM optimization problem [12]. Equation (4) can be rewrite as follows:

$$
\min_{y \in \mathbb{R}^{n_0 \times n}} \Gamma_{SS-ELM} = \frac{1}{2} \|y\|^2 + \frac{C}{2} \|y - J_{\gamma}\|^2 + \frac{\lambda}{2} T \gamma (Y^T L Y)
$$

(6)

where $L$ is Laplacian matrix. $Y$ represents the output matrix and $\gamma \in \mathbb{R}^{n_h \times n_0}$ is the output weights, respectively. $\gamma$ denotes the trade-off parameters. The parameter $C$ is a user-specified number for regularization which chose from vector set $\{10^x | x=-7, ..., 7\}$.

The gradient of equation (7) then is calculated:

$$
f_{elm}(x) = h(x)H^T \left( \frac{I}{C} + HH^T \right)^{-1} Y
$$

(7)

where $I_{n_h}$ is an identity matrix with the dimension of $n_h \times n_h$. The output weight matrix can be solved by pseudo-inverse of $H$, an alternative solution of IG-SSELM is:

$$
\gamma = H^T \left( \frac{I}{C} + HH^T \right)^{-1} Y
$$

(8)
where the $I_{[l+u]}$ is an identity matrix. If we set $\lambda=0$ in equations (8)-(9), IG-SSELM reduced to a conventional ELM.

3. Methods

In this section, we introduce the proposed IG kernel activation function and ST framework in detail, which can avoid the regularization and decision boundary which makes the model influenced by the noise in training data geometry.

3.1. IG-Kernel Function

Given the general form of activation function of ELM as ($\cdot$). Priori information plays an important role in a determined classification problem. Generally speaking, the Gaussian kernel has a satisfactory result when the data are sampled from a Gaussian Mixture distribution. However, if there is no a priori information, the activation function needs to meet the following conditions:

• $\zeta(\cdot)$ must be a nonlinear function, if $\zeta(\cdot)$ is linear, the representation capability of higher layers is not superior to the deep-layer network.

• Denote an interval boundary in $\zeta(\cdot)$ to limit the weights and activation functions.

• In the own input interval, $\zeta(\cdot)$ is a piecewise function, which means $\zeta(\cdot)$ and $\zeta'(\cdot)$ can be solved.

Define $\zeta(\cdot)$ as a monotonic function so that $\zeta'(\cdot)$ has a stable value, which can obtain the global minimum value and avoid the local optimization. S-type function generally is an activation function that can meet all the above requirements. Some common transfer functions in ELM are:

• Sigmoid kernel:

$$
\zeta(a,b,x) = \frac{1}{1+\exp(-(a \cdot x + b))}
$$

(9)

• Gaussian kernel:

$$
\zeta(a,b,x) = \exp\left(-\frac{b\|x-d\|^2}{2}\right)
$$

(10)

• Hard-limit(zero-one) kernel:

$$
\zeta(a,b,x) = \begin{cases} 1, & a \cdot x - b \geq 0 \\ 0, & a \cdot x - b \leq 0 \end{cases}
$$

(11)

A shortage of equation (11) is that only the training data geometry satisfies the corresponding Gaussian manifold that ELM can get an excellent solution. Otherwise, mathematical analysis, if the data patterns have a long distance with $a$ (or the $b$ is selected improperly big), the neuron value of Gaussian kernel activation function is closed to zero, that means the ELM hidden nodes will be very few and a lot of samples elements information is lost. To solve the issue that the output of the kernels should not be with the long tail. We propose a new activation kernel function (IG kernel function) as follows:

**Theory:** Given a kernel function $\exp(b \langle x; a \rangle)$, uniform this activation function:

$$
\frac{\exp(b(x,a))}{\sqrt{\exp(b\|x\|^2)\exp(b\|a\|^2)}} = \exp\left(b(x,a) - \frac{b(x,x)}{2} - \frac{b(a,a)}{2}\right) = \exp\left(-\frac{b\|x-d\|^2}{2}\right)
$$

(12)

which is a Gaussian function. Uniform $x$ with $x_c = \frac{x}{\|x\|^2}$ and $a_c = \frac{a}{\|a\|^2}$ in equation (12), a new activation function can be obtained:
\[
\frac{\exp(b(x, a_x))}{\sqrt{\exp(b\|x\|^2)}} = \exp\left(b(x, a_x) - \frac{b(x_x, x_x)}{2} - \frac{b(a_x, a_x)}{2}\right) = \exp\{-b(1-\cos \theta)\}
\]

where \(\langle x, a_x \rangle = \cos \theta\). Then the new activation function is:

\[
\zeta_{\text{new\_kernel}}(a, b, x) = \exp\{-b(1-\cos \theta)\}
\]

From equation (14) we can denote that the boundary of IG kernel is \([e^{-2b}, 1]\), b is a user-defined bias which commonly is set 0.5, which means \([e^{-2b}, 1]\) decrease the value of \(\text{tr}(H^TH)\) is approximately to zero in ELM, and improve the learning performance than the other activation functions.

### 3.2. The Proposed IG-SSELM Framework

In this subsection, the kernel IG function is applied to the semi-supervised ELM for reconstructing a novel model named IG-SSELM. Utilizing Mercer’s condition on ELM with a kernel activation K and replacing HHT with \(\Omega\), the solution form of equations (7-8) can be rewritten as:

\[
f_{k_{\chi \leftarrow \text{data}}} = h(x)H^T\left(\frac{I}{C} + \Omega\right)^{-1}Y = \left[K(x, x_1) \ldots K(x, x_n)\right]^T \left[\frac{I}{C} + \Omega\right]^{-1}Y
\]

And

\[
\gamma = \left[\frac{I}{C} + \Omega\right]^{-1}Y
\]

where the \(\Omega_{k,j} = K(x_k, x_j)\) \(k, j = 1, \ldots, n\). And \(h(x)H^T = [K(x, x_1), \ldots, K(x, x_n)]^T\). The procedure of IG-SSELM algorithm for this method is described as figure 2.

### 4. Experiments

To evaluate the performance of the proposed IG kernel and IG-SSELM model proposed in the paper, ST identification is utilized for a comparison with other kernel activation function ELM and other existing state-of-data ST identification methods.

#### 4.1. Experiment Setting Preparation

The industrial aluminum electrolysis reduction is given in figure 3. From expert and operator workers’ experiences, 12 variables related to SD situation are selected as input variables for model construction (see figure 4). All the experiments are conducted with MATLAB 2016a running on a 3.6-GHz i5 CPU with 32-GB RAM. To train this algorithm, 12 input process variables were normalized (-1, 1) and output of the ST label was (‘LOW’, ‘NORMAL’, ‘HIGH’). When the ST is in the range of 8 to 15, the ST is tagged as ‘NORMAL’ label; when the value of ST is 2 to 8, the ST is defined as a ‘LOW’; when the value is bigger than 15, the classification of ST is ‘HOT’. Before building the proposed method, 1200 patterns were extracted from the industrial databases, and 800 of them are the training set, and the remaining are the testing samples.
**Figure 2.** The procedure of IG-SSELM model for the SD identification.

| Input variables for ST identification |
|---------------------------------------|
| 1. The working current                |
| 2. Cell resistance                    |
| 3. Set datum time of Al₂O₃            |
| 4. Actual feeding time of Al₂O₃       |
| 5. Content of Iron                    |
| 6. Content of Silica                  |
| 7. The tapping level                  |
| 8. The aluminum electrolyte level     |
| 9. Content of aluminum                |
| 10. Electrolysis cell temperature     |
| 11. Addition amount of AlF₃           |
| 12. Average cell voltage              |

**Figure 3.** The schematic diagram of aluminum reduction cell.

4.2. **Comparison with Different Kernel Function ELM**
To demonstrate the classification performance of proposed new kernel function in ELM, linear kernel activation function (Lin-kernel ELM), polynomial kernel (Poly-kernel ELM), RBF kernel (RBF-kernel ELM) and Gaussian kernel. (Guess-kernel) is utilized as a comparison with the proposed kernel activation function. Figure 5 gives the presentation of different kernel functions. Using these kernels to ELM and the detailed results for the ST detection are described as table 1. The mathematically statistical results are presented in table 2. It illustrates that IG-kernel gets competent accuracy and relatively less training time at the same time (only a little bigger than Lin-kernel). This might because the IG kernel elevates the output weight boundary and decrease the information loss in the hidden layer.

4.3. **Comparison with Existing ST Classification Methods**
As another comparison, the IG-SSELM is also conducted on the same condition while the model is compared with existing ST identification. For example, random forest, artificial experience, and Decision Tree. All the result is introduced in table 3. The accuracy rate is competitive with other ST methods. Further, the training efficiency of IG-SSELM is superior to other methods because of the efficient matrix computation. The results show that the proposed IG-SSELM performance in aluminum production outperforms the other ST methods.
Figure 5. Comparison with different kernel activation function.

Table 1. The results of SD identification.

| ID | Predicted Label | Actual Label |
|----|-----------------|--------------|
| 1  | LOW             | NORMAL       |
| 2  | NORMAL          | LOW          |
| 3  | NORMAL          | NORMAL       |
| 4  | NORMAL          | NORMAL       |
| 5  | LOW             | LOW          |
| 6  | HIGH            | NORMAL       |
| 7  | NORMAL          | NORMAL       |
| 8  | HIGH            | HIGH         |
| 9  | NORMAL          | NORMAL       |
| 10 | LOW             | LOW          |
| 20 | HIGH            | HIGH         |

Table 2. Statistical experiment results in SD identification with different kernel activation function in ELM.

| Method    | Accuracy | Training time (s) |
|-----------|----------|-------------------|
| Lin-kernel| 0.38     | 2.14              |
| Poly-kernel| 0.42   | 3.36              |
| RBF-kernel| 0.76     | 3.88              |
| Gauss-kernel| 0.74 | 3.27              |
| IG-kernel| 0.82     | 3.01              |
Table 3. Experiment results in SD identification with different SD methods.

| Method           | Accuracy | Training time (s) |
|------------------|----------|------------------|
| ELM              | 0.42     | 2.14             |
| SS-ELM           | 0.71     | 3.36             |
| Artificial experience | 0.66   | -                |
| Random forest    | 0.74     | 2.37             |
| IG-SSELM         | 0.88     | 2.12             |

5. Conclusion

A novel kernel, RAF, is proposed in this paper. The function can avoid the influence of the outlier dataset and avoid a lot of samples that tend to 0. Further, the IG kernel is extended to semi-supervised ELM and construct an IG-SSELM to deal with the uniform distribution and scarcity of labelled data samples. Meanwhile, IG-SSELM is further applied to the ST classification problem in the aluminium electrolysis to evaluate the performances. The experiment results illustrate that IG kernel function is superior to other popular kernel activation functions. At the same time, the accuracy ST identification is competent with other existing methods. The proposed IG-SSELM framework can be extended to other industrial domains.

References

[1] Chen X F, Ying X F and Huang K K 2017 Identification of superheat of aluminium electrolytic cell based on computer vision and expert rule Chinese Automation Congress (CAC) (IEEE) pp 4705-4710.
[2] Zeng S, Li J, Wei Y, et al. 2010 Calculation and control of equivalent superheat for 300kA prebake aluminium electrolysis IEEE Intelligent Control and Automation pp 4755-4760.
[3] Chen Z, Li Y and Chen X 2017 Semantic network based on intuitionistic fuzzy directed hyper-graphs and application to aluminium electrolysis cell condition identification IEEE Access 5 20145-20156.
[4] Huang G B, Zhu Q Y and Siew C K 2006 Extreme learning machine: Theory and applications Neurocomputing 70 (1-3) 489-501.
[5] Huang G B, Zhou H, Ding X, et al. 2012 Extreme learning machine for regression and multiclass classification IEEE Transactions on Systems, Man and Cybernetics. Part B (Cybernetics) 42 (2) 513-529.
[6] Luo J, Vong C M and Wong P K 2017 Sparse Bayesian extreme learning machine for multi-classification IEEE Transactions on Neural Networks & Learning Systems 25 (4) 836-843.
[7] Wu J, Wang S and Chung F L 2011 Positive and negative fuzzy rule system, extreme learning machine and image classification Journal of Image and Graphics 2 (4) 261-271.
[8] He Q, Xin J, Du C, et al. 2014. Clustering in extreme learning machine feature space Neurocomputing 128 (5) 88-95.
[9] Belkin M, Niyogi P and Sindhwani V 2006 Manifold regularization: A geometric framework for learning from labelled and unlabelled examples Journal of Machine Learning Research 7 (1) 2399-2434.
[10] Lei Y, Chen X and Gui W 2019 Hessian regularization semi-supervised extreme learning machine for superheat identification in aluminum reduction cell Chinese Control and Decision Conference (CCDC) (IEEE) pp 4406-4411.
[11] Lei Y X, Chen X F, et al. 2020. A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification Neurocomputing 381 186-195.
[12] Rumelhart D E, Hinton G E and Williams R J 1988 Learning representations by back-propagating errors Neurocomputing: Foundations of Research (MIT Press) pp 533-536.