Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Novel Kernel-based Extreme Learning Machine with Incremental Hidden Layer Nodes

Mengcan Min * **Xiaofang Chen** ** **Yongxiang Lei** *** **Zhiwen Chen** **** **Yongfang Xie** †

* *School of Automation, Central South University, Changsha 410083, China (e-mail: mengcanmin@csu.edu.cn)*
** *School of Automation, Central South University, Changsha 410083, China (e-mail: xiaofangchen@csu.edu.cn)*
*** *School of Automation, Central South University, Changsha 410083, China (e-mail: leiyongxiang@csu.edu.cn)*
**** *School of Automation, Central South University, Changsha 410083, China (e-mail: zhiwen.chen@csu.edu.cn)*
† *School of Automation, Central South University, Changsha 410083, China (e-mail: yfxie@csu.edu.cn)*

Abstract: Extreme learning machine (ELM) is widely used in various fields because of its advantages such as short training time and good generalization performance. The input weights and bias of hidden layer of traditional ELM are generated randomly, and the number of hidden layer nodes is determined by artificial experience. Only by adjusting parameters manually can an appropriate network structure be found. This training method is complex and time-consuming, which increases the workload of workers. To solve this problem, the incremental extreme learning machine (I-ELM) is used to determine the appropriate number of hidden layer nodes and construct a compact network structure in this paper. At the same time, a new hidden layer activation function STR is proposed, which avoids the disadvantages of incomplete output information of hidden layer due to uneven distribution of sample data. The proposed algorithm is evaluated by public data sets and applied to the classification of superheat degree (SD) in aluminum electrolysis industry. The experimental results show that STR activation function has a good learning speed, and the proposed algorithm is superior to the existing SD identification algorithm in terms of accuracy and robustness.

Keywords: ELM, I-ELM, Kernel function, SD classification

1. INTRODUCTION

Single-hidden layer feedforward neural networks (SLFNs) (Gao XP et al. (2001)) is widely used in classification and regression fields because of its ability to approximate arbitrary continuous functions. However, the traditional training method of SLFNs is based on the gradient descent method (Liu YC et al. (1993)). The gradient descent method often needs multiple iterations, which has some limitations such as long training time, easy to fall into local optimal value, over-fitting the training samples and resulting in unsatisfactory test results. Aiming at the shortcomings of gradient descent method, Huang GB et al. (2006) proposed a new single-hidden layer feedforward neural network algorithm – Extreme Learning Machine (ELM). The input weights and bias of the hidden layer of the ELM are randomly generated without many iterations to find the optimal value, and remain unchanged during the training process. The number of hidden layer nodes is also manually initiated. Due to ELM has the advantages of less training parameters, fast learning speed and better generalization performance, it is widely used in various fields. For example, Zhao JX et al. (2019) proposed a convolutional neural networks (CNN) multi-layer feature fusion and extreme learning machine for diagnosis of breast disease diagnosis.Zou BX et al. (2017) applied the extreme learning machine to the research of optical fiber vibration signal recognition and obtained a good correct recognition rate. Xu YH et al. (2015) applied ELM in the field of insect classification, replacing the traditional human eye observation and recognition.

Although the application of ELM is gradually expanding, it also has some shortcomings. The traditional ELM network structure is set according to artificial experience. When the number of hidden nodes set too small, the network cannot achieve good learning effect, and when the number of hidden nodes set too much, the over-fitting phenomenon is easy to occur, which affects the generalization ability of the network. Therefore, it is often necessary to carry out many experiments and analyses to select the appropriate network structure as the final network

⋆ Sponsor and financial support acknowledgment goes here. Paper titles should be written in uppercase and lowercase letters, not all uppercase.

model. The traditional method to determine the network structure is complex and time-consuming, which increases the manual workload. Recently, in order to improve the performance of the ELM, many improved algorithms have been proposed. For example, Xu HW et al. (2018) optimized the parameters of the ELM by using the improved differential evolution algorithm and applied it in short-term load prediction, but Hu's method was greatly affected by the number of variables. Lei YX et al. (2019) introduced two different regularization methods to construct a semi-supervised ELM (SS-ELM), the SS-ELM is suitable for the case where the number of labeled samples is small and there is a large amount of unlabeled sample data. Han M (2011) proposed an ELM algorithm with 1 norm based on substitution function and Bayesian framework (N1-ELM), N1-ELM can get better test accuracy, but it has disadvantages such as large computation and long training time. Although these algorithms play a certain role in improving the performance of ELM, further research is needed on how to effectively reduce hidden layer nodes and construct appropriate network structure while ensuring learning accuracy.

In view of the complex structure of the traditional ELM, we use the I-ELM (Huang GB et al. (2006)) to determine the optimal network structure, and propose a new activation function to improve the performance of the I-ELM in this paper. The main advantages of proposed algorithm as follows:

- Use the I-ELM to determine the number of hidden layer nodes and effectively simplify the network structure.
- A new activation function is proposed to avoid uneven sample distribution and inability to learn all the information.

The experiment results demonstrate the effectiveness of the proposed method.

The rest of the paper is organized as follow. In Section 2, we give a related works introduction such as I-ELM and related activation function, which lays the foundation of the proposed method. In Section 3, we propose the STR kernel function and introduce the proposed STR-IELM algorithm. The public dataset experimental results and SD classification in industrial aluminum electrolysis process are presented in Section 4. Finally, Section 5 concludes the paper.

2. RELATED WORK

In this section, we mainly give a brief introduction of I-ELM, and related kernel function used in ELM, which lays the foundation of our later research.

2.1 I-ELM

In view of the traditional ELM network structure is set up by artificial experience, the reference (Romero E (2002)) proposed the incremental extreme learning machine, whose network structure is shown in Fig. 1.

When the I-ELM starts training, the network structure does not exist, that is, the number of nodes in the hidden layer is set to 0. The network training error is reduced

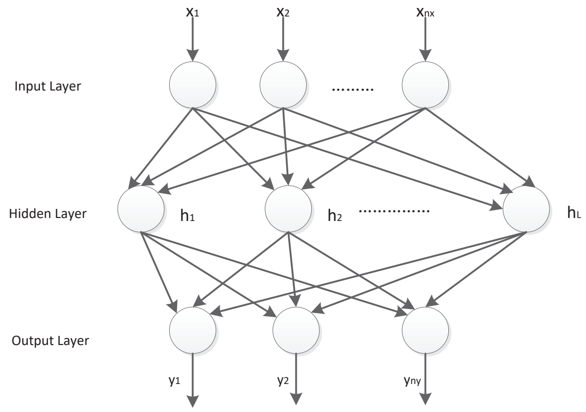

Fig. 1. The framework of I-ELM

by adding hidden layer nodes to the network continuously (Huang GB et al. (2007)). The training process will not stop until the error is less than the expected error ε or the number of hidden layer nodes reaches the set maximum value $L_{\max}$. In I-ELM theory, the output weight β based on the training data sets is:

$$\beta_{\mathrm{n}}=\frac{E \cdot H^T}{H \cdot H^T} \tag{1}$$

where $\mathbf{E}$ is the residual vector, that is, the difference between the actual output of the network and the expected output, whose initial value is output $\mathbf{T}$. The maximum number of neurons in the hidden layer is $L_{\max}$, and the expected learning accuracy is ε. The number of hidden neural nodes increases from 0.

The proof process of the output weight value Eq. (1) is as follows:

Proof: Assume that the network has only one output, so $\mathbf{E}=[\mathrm{e}_1,\mathrm{e}_{2,\ldots,}\mathrm{e}_n]$, $\mathbf{H}=[h_1,h_2,...,h_n]$

(1)According to Eq. (1), the output weight β is:

$$\beta=\frac{E \cdot H^T}{H \cdot H^T}=\frac{\sum\limits_{\mathrm{i}=1}^{N} e_i h_i}{\sum\limits_{\mathrm{i}=1}^{N} h^2} \tag{2}$$

(2)Calculating the residual error:

$$\mathbf{E}-\beta\mathbf{H}=[(e_1-\beta h_1),(e_2-\beta h_2),...,(e_N-\beta h_N)] \tag{3}$$

(3)According eq (3) to calculation error Q:

$$\begin{aligned} Q &= \sqrt{\frac{1}{N}\sum_{\mathrm{i}=1}^{N}(e_i-\beta h_i)^2} \\ &= \sqrt{\frac{1}{N}(\sum_{\mathrm{i}=1}^{N} e_i^2 - 2\sum_{\mathrm{i}=1}^{N} e_i h_i \beta + \sum_{\mathrm{i}=1}^{N} h_i^2\beta^2)} \end{aligned} \tag{4}$$

(4)Find the weight β that minimizes the error Q:

$$\zeta(\beta)=\sum_{\mathrm{i}=1}^{N} e_i^2 - 2\sum_{\mathrm{i}=1}^{N} e_i h_i\beta + \sum_{\mathrm{i}=1}^{N} h_i^2\beta^2 \tag{5}$$

Take the derivative of $\zeta(\beta)$:

$$\dot{\zeta}(\beta)=2\sum_{\mathrm{i}=1}^{N} h_i^2\beta - 2\sum_{\mathrm{i}=1}^{N} e_i h_i \tag{6}$$

By solving Eq. (6), the expression of output weight can be obtained by Eq. (2).

2.2 Several Activation function

In the ELM, the choice of activation function has great influence on the performance of the algorithm. We can choose the appropriate activation function according to the priori information of the data. For example, a satisfactory result can be obtained by selecting the gaussian activation function when the data sample fits the gaussian mixture distribution. However, it is a complex and time-consuming work to analyze the data and extract the distribution information. According to the reference (Richard O. Duda (2000)), the algorithm can also achieve good results when the activation function selected satisfies the following conditions:

(1) $\mathbf{J}$ $(\cdot)$ must be nonlinear, if $\mathbf{J}$ $(\cdot)$ is linear, the learning ability of network with multiple hidden layers is not higher than the neural network with two hidden layers.

(2) The output value of $\mathbf{J}$ $(\cdot)$ must have a maximum and a minimum value, which is used to limit the boundary of weights and activation functions.

(3) $\mathbf{J}$ $(\cdot)$ must be continuous and smooth, which means $\mathbf{J}$ $(\cdot)$ and $\mathbf{J}$' $(\cdot)$ ($\mathbf{J}$' $(\cdot)$ represents the derivative of the $\mathbf{J}$ $(\cdot)$ function) have definitions under their own independent variables ranges.

Some popular activation functions (Lohani HK et al. (2019)) used in ELM are listed as follows:

- Sigmoid function:

$$\mathbf{J}(\mathbf{a},\mathbf{b},\mathbf{x}) = \frac{1}{1+\exp(-(\mathbf{ax}+\mathbf{b})} \tag{7}$$

- Gaussian function:

$$\mathbf{J}(\mathbf{a},\mathbf{b},\mathbf{x}) = \exp(-\frac{\mathbf{b}\|\mathbf{x}-\mathbf{a}\|^2}{2}) \tag{8}$$

- Hard-limit(zero-one) function:

$$\mathbf{J}(\mathbf{a},\mathbf{b},\mathbf{x}) = \begin{cases} 1 & \mathbf{ax}-\mathbf{b} \geq 0 \\ 0 & \mathbf{ax}-\mathbf{b} \leq 0 \end{cases} \tag{9}$$

The intuitive geometric description of Sigmoid function, Gaussian function, Hard-limit functions is shown in the Fig. 2.

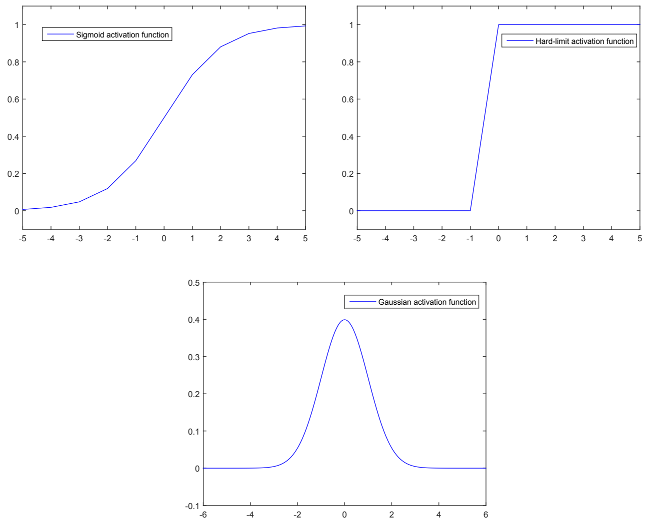

Fig. 2. The curve of different activation function, (a) Sigmoid activation function (b) Hard-limit activation function (c) Gaussian activation function

3. PROPOSED FRAMEWORK

3.1 A new kernel function STR

Due to the random acquisition of network input weights and hidden layer neuron bias, the hidden layer output $\mathbf{H}$ is also random, and the residual vector $\mathbf{E}$ is given. As you can see from Fig. 2, these activation functions all have regions tending to zero. Therefore, when output weight β is obtained according to Eq. (2), $\sum_{i=1}^{N} e_i h_i$ may approach to zero, so that the output weight β obtained is too small, and the hidden layer node is invalid.

In order to avoid the hidden layer output $\mathbf{H}$ close to zero, we proposed a new activation kernel function—STR kernel function:

$$\mathbf{J}(\mathbf{x}) = \begin{cases} \frac{\mathbf{e^x}-\mathbf{e^{-x}}}{\mathbf{e^x}+\mathbf{e^{-x}}} & \mathbf{x} \leq 0 \\ \mathbf{x} & 0 < \mathbf{x} < 1 \\ 1 & 1 \leq \mathbf{x} \end{cases} \tag{10}$$

The intuitive geometric description of STR activation functions is shown in the Fig. 3. The boundary of STR kernel function is [-1,1]. The STR function avoids the disadvantage that the output value $\mathbf{H}$ of the hidden neuron tends to zero, which leads to the output weight β being too small. It introduces negative activation and increases the generalization ability of the network. Compared with other activation functions, the STR activation function improves the performance of ELM.

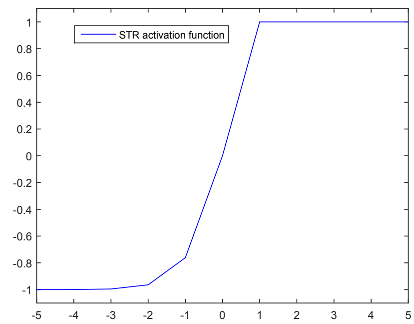

Fig. 3. The curve of STR activation function.

3.2 The model based on STR

The detail of STR-IELM algorithm procedure is stated in **Algorithm 1**. The training process will stop only when the number of hidden nodes L exceeds the preset maximum number $L_{\max}$ or the residual error $\mathbf{E}$ of the current network structure is less than the expected error ε. In Eq. (13), $\mathbf{E}$ represents the residual error vector before the new node added and $\mathbf{E}^*$ represents the residual error vector after the new node added.

Algorithm 1 STR-IELM Algorithm

Input: Training data sets $\{(\mathbf{x}_j,\mathbf{t}_j)\}_{j=1}^{N}$; Activation function J(x); Maximum node number $L_{\max}$; Expect error ε.

Output: The output weight matrix β_n for I-ELM

step 1: Whether $L < L_{\max}$ and $\|\mathbf{E}\| > \varepsilon$

step 2: $L = L + 1$

step 3: Initiate an ELM framework of L hidden neurons with random weights **a** and biases **b**:

$$\mathbf{a}^{\mathbf{T}}\mathbf{a} = \mathbf{I} \quad \mathbf{b}^{\mathbf{T}}\mathbf{b} = 1 \tag{11}$$

step 4: Calculate the input $\mathbf{x}$ of the current activation function: $\mathbf{x} = \mathbf{a}\mathbf{x} + \mathbf{b}$

step 5: Using Eq. (10) to calculate the output of the hidden layer $\mathbf{H} = \mathbf{J}(\mathbf{x})$

step 6: Calculate the output weights of hidden layer neurons:

$$\beta_n = \frac{E \cdot H^T}{H \cdot H^T} \tag{12}$$

step 7: Calculate the residual error after adding the new hidden node L:

$$\mathbf{E}^* = \mathbf{E} - \beta\mathbf{H} \tag{13}$$

step 8: Back to step 1

At the beginning of training, the maximum number of hidden layer nodes is set to 100. When the residual error **E** of the trained network structure is less than the set expected error ε, the network structure at this time is the most suitable. Otherwise, training is started from STR-IELM with 100 hidden layer nodes, and the maximum hidden layer node number is set to 200. Increase the number of hidden layer nodes in turn until the most suitable network structure is found. The proposed STR-IELM structure framework is showed in Fig. 4.

4. EXPERIMENT RESULTS

In this section, to verify the performance of our proposed STR activation, we conducted several comparative experiments on four public datasets. At the same time, the traditional ELM structure and the I-ELM network structure were compared. Finally, in order to calculate the performance of the proposed STR-IELM method, it was applied to the aluminum electrolysis industry process, and several existing SD identification methods are compared (see for examples Yue WC et al. (2017); Gui WH et al. (2018)). All experimental results are given in this section. The experiment environment is based on MATLAB 2016a running on a 3.6-GHz i5 CPU with 16-GB RAM.

4.1 Experiment results on public datasets

All experimental data are from UCI database repository [1], and the range of data after normalization is (-1,1). Four kinds of experimental data are described in detail in Table 1. 50 % of each group of data was randomly selected as the training datasets, and the rest was used as the testing datasets.

To evaluate the performance of STR kernel activation function proposed in this paper, we compared the I-ELM

[1] http://archieve.ics.uci.edu/ml/

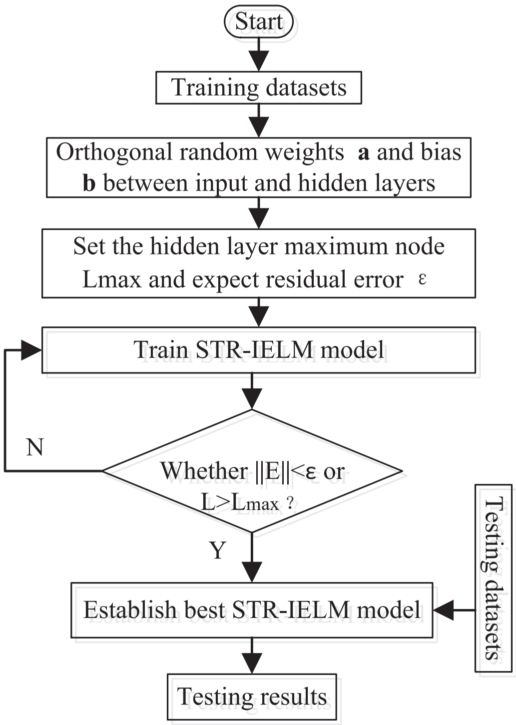

Fig. 4. The flowchart of algorithms

with Sigmoid activation function (S-IELM), I-ELM with Hard-limit activation function (H-IELM), I-ELM with Gaussian function (G-IELM), I-ELM with STR activation function (STR-IELM). The experimental results: training time and recognition accuracy are given in Table 2, Table 3, respectively. The experimental results were obtained by averaging 10 training sessions. Whether in the banknote authentication data set or the other three data sets, the accuracy rate of STR-IELM is higher than that of S-IELM, H-IELM and G-IELM, and the training time of STR-IELM is shorter. Experimental results show that STR-IELM has better learning performance than S-IELM,H-IELM and G-IELM under the same condition.

Fig. 5 shows the rising curve of recognition accuracy of S-IELM, H-IELM, G-IELM and STR-IELM when I-ELM network structure is added to 100 neurons in hidden layer. It can be seen from the Fig. 5 the accuracy rate of STR-IELM rises faster than that of I-ELM algorithm using the other three activation functions, mainly because STR activation functions have faster learning speed.

Table 1. Description of experimental data

Name	Training data	Testing data	Attribute	classes
Banknote authentication	686	686	5	2
Balance scale	313	312	5	3
Glass	107	107	10	7
iris	75	75	5	3

We've also compared the traditional ELM and I-ELM. Here, RMSE index is used to evaluate the performance of the proposed method, and its calculation formula is :

Table 2. Training time (s) of I-ELM with different kernel activation function

Dataset	Sigmoid	Hard-limit	Gauss	STR-Kernel
Banknote authentication	0.93547	1.0102	1.2253	0.90203
Balance scale	0.75484	0.78344	0.81219	0.71109
Glass	0.67797	0.76703	0.79469	0.64063
Iris	0.54094	0.61031	0.64812	0.51922

Table 3. Accuracy of I-ELM with different kernel activation function

Dataset	Sigmoid	Hard-limit	Gauss	STR-Kernel
Banknote authentication	86.573	84.283	81.927	90.278
Balance scale	91.417	87.692	89.256	94.118
Glass	68.857	65.469	67.734	75.107
Iris	90.667	88.773	89.72	93.75

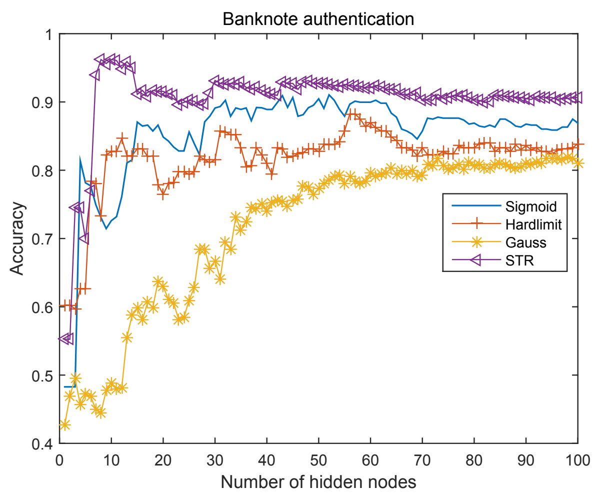

Fig. 5. Comparison I-ELM network for different activation functions on Banknote authentication data.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t-\tilde{t})^{2}} \quad (14)$$

Where $\tilde{t}$ presents the actual output value and t is the expect output value, N denotes the number of testing samples. All the experimental results are given in Table 4 and Table 5, the results show that I-ELM has higher identification accuracy than ELM, it mainly because I-ELM has more compact network structure, as well as stronger learning and generalization ability.

4.2 Comparison with existing SD classification methods

SD is an important index to evaluate the production status in the process of aluminum electrolysis industry, it refers to the difference between the initial crystal temperature of the electrolytic cell and the electrolyte temperature. The traditional method to get SD value is that workers use instruments to measure the electrolyte temperature, whereas an aluminum electrolysis plant is a high-temperature,

Table 4. Accuracy of different ELM with different activation function

Dataset	S-ELM	STR-ELM	S-IELM	STR-IELM
Banknote authentication	72.012	75.182	86.573	90.278
Balance scale	81.41	85.256	91.417	94.118
Glass	55.981	63.593	68.851	75.107
Iris	83.57	89.027	90.667	93.75

Table 5. Testing RMSE of different ELM with different activation function

Dataset	S-ELM	STR-ELM	S-IELM	STR-IELM
Banknote authentication	0.0953	0.0858	0.0726	0.0612
Balance scale	0.0781	0.0689	0.0548	0.0584
Glass	0.2589	0.1164	0.1027	0.0892
Iris	0.0748	0.0671	0.0605	0.0569

corrosive gas containing environment. Affected by the environment, the measuring instrument is easy to wear out and the accuracy of measuring result is low. Meanwhile, there will be errors in the manual reading. Therefore, how to accurately obtain SD and reduce the economic cost is a major challenge in the process of aluminum electrolysis industry. According to technicians and experts operation experience, industrial process variables that can directly affect SD were selected as input variables (Lei Y et al. (2019, 2020)). The specific details of these input variables are shown in Table 6. Before training algorithm, all the input variables are normalized to (-1,1).

The experiment collected 500 samples aluminum electrolysis database, of which 300 are training data sets and the remaining 200 are test training sets. We compared the identification methods commonly used for SD in the aluminum electrolytic industry process, such as SVM (Wang DM et al. (2015)), Random forest (Liu YS et al. (2017)), Artificial experience, ELM. The details of experimental results are given in Table 7. The experimental results show under the same experimental condition, the training time of STR-IELM is longer than the ELM with sigmoid activation function, which is mainly because it takes STR-IELM some time to determine the number of hidden layer nodes in the training process. However, among the five algorithms, STR-IELM has the highest recognition accuracy rate, and the accuracy rate based on artificial experience is the lowest compared with the other four algorithms. Therefore, it can be obtained that STR-IELM has better network performance.

5. CONCLUSION

In this paper, we proposed a new activation kernel function and apply it to the ELM. Four groups of public data sets are selected to compare STR activation function with several other commonly used activation functions of ELM, the experimental results show that STR activation function performs better than other activation functions. Further, in order to avoid the randomness of generating the number of hidden layer nodes in the ELM, we used incremental algorithm to construct STR-IELM. However,

Table 6. Selected process variables according operator experiences

No.	Variables	Descriptions
1	x_1	Current working current
2	x_2	Bath resistance
3	x_3	The content of Iron
4	x_4	The content of Silica
5	x_5	The electrolyte level
6	x_6	Molecular ratio
7	x_7	Average bath voltage
8	x_8	Addition amount of AlF_3

Table 7. Experiment results with different methods

Methods	Accuracy	Training times
SVM	0.726	1.85
Random forest	0.784	2.1
Artificial experience	0.631	–
ELM	0.695	0.436
STR-IELM	0.863	0.859

there are some problems need to be further investigated. Every time an additional hidden node is added, the output weight of the hidden layer needs to be recalculated, so the training time of the STR-IELM network model is longer than that of the traditional ELM. STR-IELM is applied to the identification of SD in the electrolytic aluminium industry, and compared with several traditional SD identification algorithms such as ELM, SVM and Random forest, the experimental results show that STR-IELM has higher network accuracy and more stable network structure. The proposed STR-IELM network framework can be further applied to other classification and regression problems and extended to more application areas.

ACKNOWLEDGEMENTS

The research is supported by the National Natural Science Foundation of China (No.61773405, No.61751312), the Major Scientific and Technological Innovation Projects of Shandong Province(2019JZZY020123).

REFERENCES

Gao XP, and Cong S (2001). Comparative Study on Fast Learning Algorithms of BP Networks. *Control and decision*,16(02):167–171.

Liu YC, and Zhang JY (1993). Gradient Descent Method. *Journal of East China Institute of Technology*, 02: pages: 12-16+22.

Huang GB, Zhou H, Ding X, et al. (2006). Extreme learning machine:theory and applications. *Neurocomputing*,70 (1-3): 489–501.

Zhao JX, Qian YR, Nan FZ, et al. (2019). The method with CNN multilayer feature fusion and ELM diagnosis for breast diseases. *Computer Engineering and Applications*, 1–7.

Zou BX, Miao J, Xu SW, et al. (2017). Research on vibration signal recognition of optical fiber based on ELM algorithm. *Computer Engineering and Applications*, 53(16): 126–133.

Xu YH, Qi HF (2015). Insert Classification Based on the Theory of the ELM. *Electronic Sci.&Tech*, 28(03): 33–37.

Hu HW, Yang Y, Shi W, et al. (2018). A short-term load prediction method based on ELM. *Heilongjiang Electric Power*, 40(06): 471–476.

Lei YX, Cen LH, Chen XF, Xie YF (2019). A Hybrid Regularization Semi-Supervised Extreme Learning Machine method and Its Application. *IEEE Access*, 07: 430102-30111.

Han M, Li DC (2011). An norm 1 regularization term ELM algorithm based on surrogate function and bayesian framework. *Acta Automatic Sinica*, 37(11): 1344–1350.

Huang GB, Lei C, Siew CK (2006). Universal approximation using incremental constructive feedforward network with random hidden nodes. *IEEE Transactions on Neural Networks*, 17(4): 879–892.

Romero, E., Alquezar, R. (n.d.) (2002). A new incremental method for function approximation using feed-forward neural networks. *In: International Joint Conference on Neural Networks*, 2: 1968–1973.

Huang GB, Li MB, Lei C, Siew CK (2007). Incremental extreme learning machine with fully complex intermediate nodes *Neurocomputing*, 71(4-6): 576–583.

Richard O. Duda. Peter E. Hart. David G. Stork (2000). *Pattern Classification*, 2nd ed. Wiley-Interscience, New York.

Lohani HK, Dhanalakshmi S, Hemalatha V (2019). Performance Analysis of Extreme Learning Machine Variants with Varying Intermediate Nodes and Different Activation Functions. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. *Advances in Intelligent Systems and Computin*, vol 768. Springer, Singapore.

Yue WC, Chen XF, Gui WH, et al. (2017). A knowledge reasoning fuzzy-bayesian network for root cause analysis of abnormal aluminum electrolysis cell condition. *Front Chem Sci Engineer*, 11(3): 414–428.

Gui WH, Yue WC, Chen XF (2018). Process industry knowledge automation and applications in aluminum reduction production process. *Control Theory Control Appl*, 35(7): 887–899.

Chen ZG, Li YG, Chen XF, et al. (2017). Semantic Network Based on Intuitionistic Fuzzy Directed Hyper-Graphs and Application to Aluminum Electrolysis Cell Condition Identification. *IEEE Access*, 5: 20145–20156.

Lei Y, Chen X, Min M, et al. (2020). A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification. *Neurocomputing*, 381: 186-195.

Lei Y, Chen X, Gui W (2019). Hessian Regularization Semi-supervised Extreme Learning Machine for Superheat Identification in Aluminum Reduction Cell. *2019 Chinese Control And Decision Conference (CCDC). IEEE*, 4406-4411.

Wang DM, Lu CH, Jiang WW, Li BR (2015). Study on PSO-based decision- tree SVM multi-class classification method. *Journal of electronic measurement and instrumentation*, 29(04): 211–615.

Liu YS, Xia SY, Yu H, et al. (2017). Prediction of Aluminum Electrolysis Superheat Based on Relative Density Noise Filtering Random Forest. *Chinese Automation Congress(CAC)*.