An Efficient Extreme Learning Machine Based on Fuzzy Information Granulation

http://dx.doi.org/10.3991/ijoe.v11i8.4884

Xia-fu LV, Jun-peng CHEN, Lei LIU, Bo-hua WANG, and Yong WANG
Chongqing University of Posts and Telecom, Chongqing, China

***Abstract*—To improve learning efficiency and generalization regarding extreme learning machines (ELM), an efficient extreme learning machine based on fuzzy information granulation (FIG) is put forward. First, using FIG to eliminate redundant information in the original data set and then ELM to train granulated data for prediction, this method not only improves the speed of basic ELM algorithm that contains many hidden nodes, but also overcomes the weakness of by getting rid of redundant information in the observed values. The experimental results show that the proposed method is effective and can produce desirable generalization performance in most cases based on a few regression and classification problems.**

***Index Terms*—Extreme learning machine (ELM), Fuzzy information granulation (FIG), Neural networks, Support vector machine (SVM)**

I. INTRODUCTION

An extreme learning machine (ELM), proposed by Dr. Huang in 2004, works for the "generalized" single-hidden layer feed-forward networks (SLFNs) [1,2], but there is no need to tune the hidden layer (called feature mapping) in the ELM [3]. Because ELM's input weights and hidden neurons' biases do not need to be adjusted during training, and one may randomly assign values to them, the learning phase of many applications can be completed within seconds [4]. Compared to some classical methods, such as neural networks [5] and support vector regressions (SVR) [6], ELM requires fewer optimization constraints and results in simpler implementation, faster learning, and better generalization performance.

Due to the simplicity of ELM's implementations, ELM has been extensively used in classification and regression applications. However, because of some ELM's parameters (the input weights and hidden biases) are randomly chosen, ELM's learning efficiency and generalization ability cannot be guaranteed. Usually, there are two ways to solve this problem. One is to find the best parameters; some scholars have used intelligence algorithms. The other is to increase the number of the hidden units to make these randomly generated parameters approach the best parameters [7]. Nevertheless, an increase in the hidden units will add to the amount of calculation and is inconvenient for implementation.

In recent years, information granulation (IG) has been increasingly used as an effective technique to get rid of redundant information in the observed values. IG is the process of forming meaningful entities of information, and fuzzy modeling can be conveniently adopted for information granulation, i.e. fuzzy information granulation (FIG). The FIG approach can transform primary data into a sequence of granules by setting the size of the granulation window, generating granulated sets [8]. Historical observation data are selected to be fuzzy information granulated, and this process can improve training time and training efficiency and ensure test accuracy.

Enlightened by the idea of a FIG-SVM (support vector machine based on fuzzy information granulation) algorithm [9], this paper puts forward an efficient extreme learning machine (ELM) based on FIG, called FIG-ELM. The FIG-ELM was able to achieve good generalization performance and prediction accuracy with less training time. In this paper, two regression problems and two classification problems are used to examine FIG-ELM's performance, and the results show that the FIG-ELM algorithm is superior to conventional ELM, SVM, FIG-SVM, BP, LVQ and DT [10].

II. EXTREME LEARNING MACHINE

The typical single hidden layer feed-forward networks (SLFNs) can be expressed as:

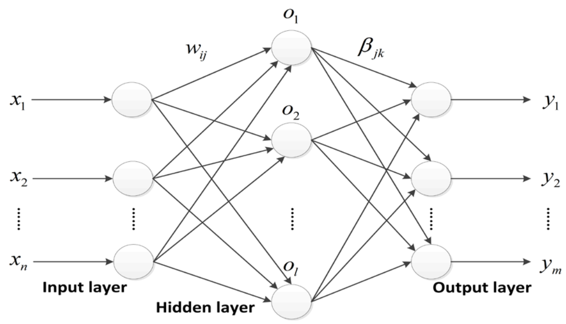


Figure 1. SLFN structure

Where x and y, respectively, represent the input element and the output element, Wij is the weight vector connecting the i-th input node and the j-th hidden node, β_{jk} is the weight vector connecting the j-th hidden node and the k-th output node.

For Q, arbitrary distinct samples (x_i, y_i), $x_i=[x_{i1},x_{i2},L,x_{in}]^T \in R^n$, $y_i=[y_{i1},y_{i2},L,y_n]^T \in R^m$, i=1,2,k,Q, g(x) are the activation function and b is the threshold. The ELM's mathematic model is:

$$T=\sum_{i=1}^{l}\beta_i g_i(x_j)=\sum_{i=1}^{l}\beta_i g(w_i \cdot x_j+b_i),\ j=1,2,K\ ,Q \tag{1}$$

where $w_i=[w_{i1}, w_{i2}, K, w_{in}]^T$ is the weight vector connecting the i-th input node and the hidden nodes, $x_j=[x_{1j}, x_{2j}, K, x_{nj}]^T$ is the input matrix, $\beta_i=[\beta_{i1}, \beta_{i2}, K, \beta_{in}]^T$ is the weight vector connecting the i-th hidden node and the output nodes, and bi is the threshold of the i-th node, i=1, 2, k, l ..

Eq. (1) can be rewritten as follows:

$$H\beta = T' \qquad (2)$$

where T' is the transposition of matrix T, and H is called the hidden layer output matrix in the neural network.

$$H(w_1, w_2, K\ , w_l, b_1, b_2, K\ , b_l, x_1, x_2, K\ , x_Q,) =$$

$$\begin{bmatrix} g(w_1 \cdot x_1 + b_1) g(w_2 \cdot x_1 + b_2) g(w_l \cdot x_1 + b_l) \\ g(w_1 \cdot x_2 + b_1) g(w_2 \cdot x_2 + b_2) g(w_l \cdot x_2 + b_l) \\ M \\ g(w_1 \cdot x_Q + b_1) g(w_2 \cdot x_Q + b_2) g(w_l \cdot x_Q + b_l) \end{bmatrix}_{Q*l} \qquad (3)$$

According to (1) and (2), matrix H is known when activation function g(x) is given and the parameter values (w, b) are randomly obtained. So the weight vector β can be solved by the following equation:

$$\min_{\beta} \| H\beta - T' \| \qquad (4)$$

The solution is:

$$\hat{\beta} = H^{+} T' \qquad (5)$$

where H+ is matrix H's Moore-Penrose generalized inverse.

III. AN EFFICIENT EXTREME LEARNING MACHINE BASED ON FUZZY INFORMATION GRANULATION (FIG-ELM)

A. *Fuzzy Information Granulation (FIG)*

The concept of FIG was suggested by Dr. Lotfi and A. Zadehin in the 1960s. The FIG approach was implemented to transform the original data into a sequence of granules, gaining a more general view at the data that retains only the most dominant components of the original temporal series [8].

For a given time series, all time series X can be regarded as a window for fuzzification. The task of fuzzification is to create a fuzzy granule P on X, which can reasonably describe an inkling G of X [9]. So the definition of data are:

g @x is G (6)

Fuzzification essentially is a process to ensure a function A, in which A is the membership function of G. Generally speaking, there are some common forms of fuzzy granules: triangular type, trapezoidal type, Gaussian type, parabolic type, and so on. In this paper, a triangular type was chosen. The membership function can be constructed as:

$$A(x,a,m,b) = \begin{cases} 0, x < a \\ \dfrac{x-a}{m-a}, a \le x \le m \\ \dfrac{b-x}{b-m}, m < x \le b \\ 0, x > b \end{cases} \qquad (7)$$

The structure of a triangular type is shown in Figure 2.

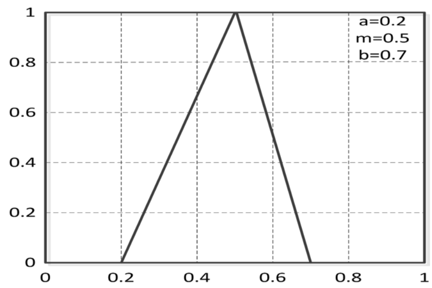


Figure 2. Triangular type structure

B. *FIG-ELM*

ELM need not spend much time tuning the input weights and hidden biases of the SLFN by randomly choosing these parameters. Since the output weights are computed based on the input weights and hidden biases, there inevitably exists a set of non-optimal or unnecessary input weights and hidden biases. Thus, to make these randomly generated parameters approach the best parameters, we need to increase the number of the hidden units at the cost of increasing the training time.

In view of the above situation, an efficient approach named FIG-ELM, combining ELM with FIG, was proposed. To improve ELM's training time, learning efficiency, and generalization ability, FIG was exploited to dispose of the original data set, namely to get rid of redundant information in the observed values, and then ELM was used to do train granulated data for prediction.

The whole arithmetic processes is shown in Figure 3:

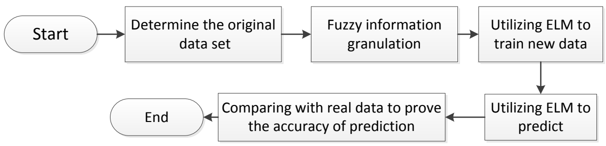


Figure 3. Flow chart

IV. EXPERIMENTAL RESULTS AND ANALYSES

To prove the capabilities of the approach proposed in this paper, some regression problems and classification problems were selected. All the programs were run in a MATLAB 2010a environment. In every simulation, there were three different fuzzy particles after granulation: Low, R, and Up. We only choose fuzzy particle Low for prediction to facilitate the process.

A. *Regression Problems*

a. Real-World Problem

The performances of SVM, FIG-SVM, ELM, and FIG-ELM were compared on a real-world benchmark data set (Shanghai composite index). We chose 400 groups of data.

There were 100 hidden nodes assigned for ELM, FIG-ELM and SVM. FIG-SVM's parameters were C=10 and g=1.5. Figure 4 and Figure 5 show the original data and the new data disposed by FIG, respectively. Figure 6 shows the fitting results obtained by these four learning algorithms. Table 1 summarizes the results of the real-world regression problem with regard to the training time, maximum absolute error, least absolute error, and average error for making the proposed algorithm's performance prominent.

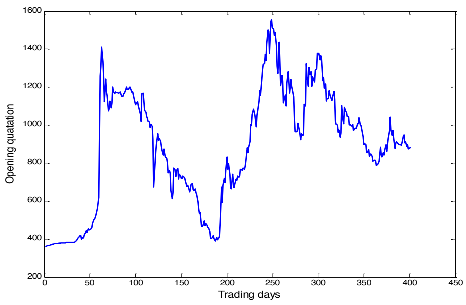


Figure 4. Original data

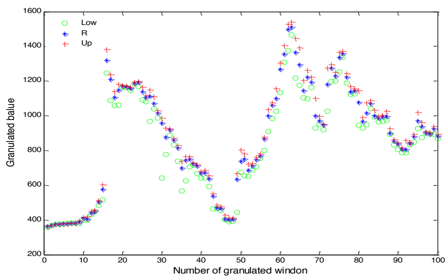
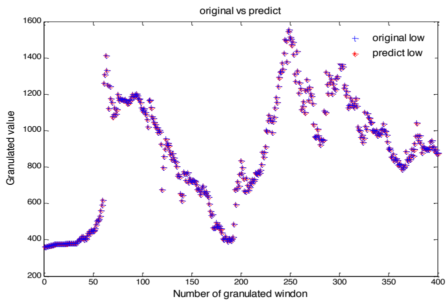
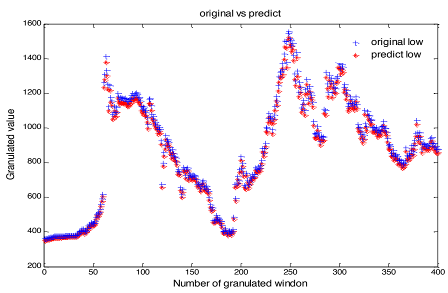


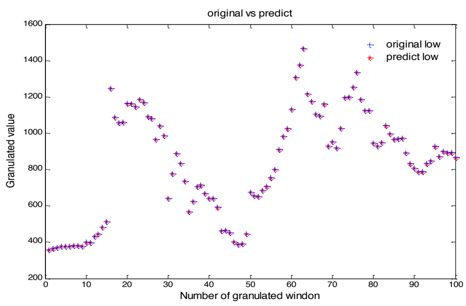
Figure 5. Granulation data



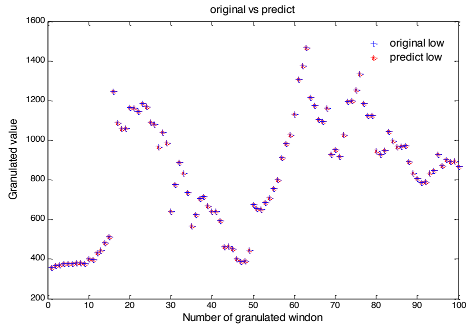
(a) SVM



(b) ELM



(c) FIG-SVM



(d) FIG-ELM

Figure 6. Prediction performances of four algorithms

TABLE I.
PERFORMANCE COMPARISON BETWEEN FOUR ALGORITHMS

Algorithms	Train Time (s)	Maximum Absolute Error	Least Absolute Error	Average Absolute Error
SVM	2.39	0.98	0.02	0.55
FIG-SVM	1.57	0.96	0.05	0.43
ELM	0.73	36	7	21
FIG-ELM	0.45	5.2*10-7	1.1*10-7	3.3*10-7

It can be seen from Table I that the FIG-ELM took the least time and had the optimal performance. The FIG-ELM learning algorithm spent 0.45s obtaining the average absolute error ($3.3*10^{-7}$). However, it respectively took 2.39s, 1.57s and 0.73s for SVM, FIG-SVM and ELM algorithms to reach a much higher average absolute error (0.55, 0.43 and 21). In addition, it is clear in Figure 6 that the prediction results obtained by FIG-ELM almost coincide with the true values. Obviously, it can be seen that the proposed algorithm outperformed the contrastive algorithms.

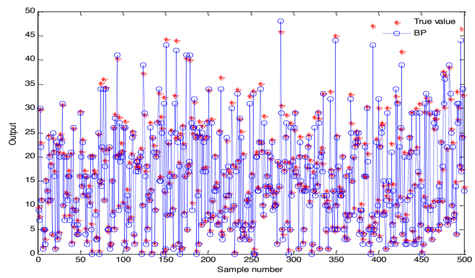
b. Approximation of nonlinear function

In this example, four algorithms (BP, SVM, ELM, and FIG-ELM) were used to approximate the nonlinear function. The nonlinear function is represented as:

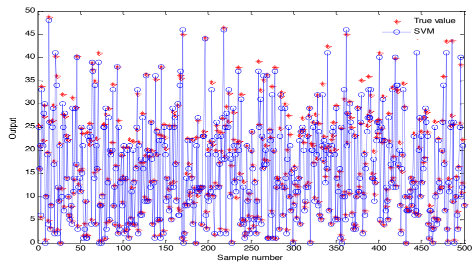
$$y=x_1^2+x_2^2 \quad (8)$$

The training data set and testing data set were randomly generated, and 1,500 groups of data were used for training and 500 groups for testing.

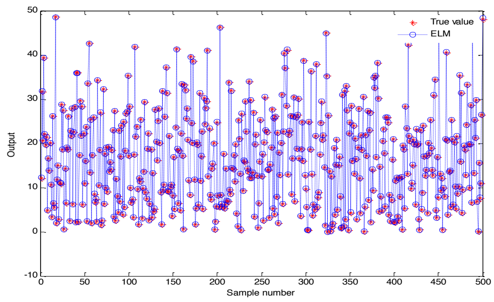
The 100 hidden nodes assigned for BP, ELM, FIG-ELM and SVM's parameters are C=10 and g=1.5. Moreover, FIG's number of granulation windows is 300, namely FIG-ELM's 300 groups of training data can represent 1,500 groups. The training time, average absolute error, and testing mean square error (MSE) of these four algorithms are shown in Table II. Figure 7 shows the true value of the nonlinear function and the approximated value obtained by these four learning algorithms.



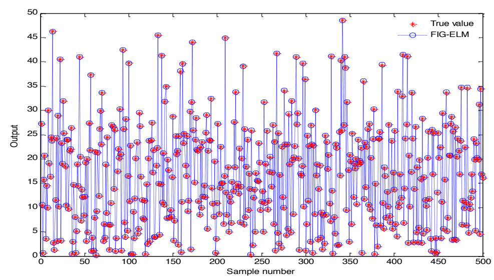
(a) BP



(b) SVM



(c) ELM



(d) FIG-ELM

Figure 7. Outputs of BP, SVM, ELM and FIG-ELM learning algorithms

TABLE II.
PERFORMANCE COMPARISON OF NONLINEAR FUNCTION

Algorithms	Train Time (s)	Average Absolute Error	MSE
BP	2.07	0.467	0.66
SVM	0.62	0.552	0.88
ELM	0.045	2.5*10-3	3.7*10-3
FIG-ELM	0.02	1.45*10-4	5.7*10-4

As can be seen from Figures 7(a) and (b), the BP and SVM learning algorithms' testing error is obvious. We can see from Figure 7 and Table II that FIG-ELM has the best prediction performance; it not only has the shortest training time but also the minimum MSE.

B. Classification Problems

a. Small-scale Real Classification Application

Performance comparisons of the proposed FIG-ELM algorithm and many other popular algorithms (BP, SVM, LVQ and DT) were conducted through a real medical diagnosis problem (breast cancer) produced at the University of Wisconsin-Madison. There were 10 attributes in total, and the hidden nodes were set as 100. The total data were 569 groups; 500 groups of data were randomly selected for training, and the other 69 groups for testing. Similarly, FIG's number of granulation windows was 100, namely the FIG-ELM's 100 groups of training data represented 500 groups. One hundred trials were conducted to compare the performance of these algorithms. Simulation results, including the average training accuracy, the average testing accuracy, and the training time, are shown in Table III.

TABLE III.
PERFORMANCE COMPARISON IN REAL MEDICAL DIAGNOSIS APPLICATION: BREAST CANCER

Algorithms	Training Time (s)	Training Accuracy (%)	Testing Accuracy (%)
BP	2.06	81.16	73.9
SVM	1.03	100	67.97
LVQ	0.33	89	91.33
DT	0.25	87.85	90
FIG-ELM	0.07	95.74	95.65

It is easy to see from Table III that SVM's testing accuracy is lowest, although it obtained the best training accuracy. BP performed very poor in this case, with its testing accuracy at 0.739 with the longest training time. LVQ and DT algorithms obtained better performance at training time and training and testing accuracy, but was still not good. By contrast, the FIG-ELM algorithm ran the fastest among the five algorithms and obtained the best performance.

b. Medium Size Classification Problem—Character Recognition

The total data were 6,000 groups; 5,000 groups of data were randomly selected for training, and the other 1,000 for testing. In addition, the FIG's number of granulation windows was 500. One hundred trials were conducted to compare the performance of these algorithms. Simulation results, including the average training accuracy, the average testing accuracy, the training time and the number of hidden nodes are shown in Table IV.

TABLE IV.
PERFORMANCE COMPARISON IN MEDIUM SIZE APPLICATIONS: CHARACTER RECOGNITION

Algorithms	Hidden Nodes	Training Time (s)	Training Accuracy (%)	Testing Accuracy (%)
SVM	-	77.2	100	86.6
ELM	50	0.22	81.64	81.2
FIG-ELM	50	0.13	93.78	91.4
ELM	100	1.13	90.54	89.2
FIG-ELM	100	0.16	97.58	92.7
ELM	500	2.04	93.62	91.3
FIG-ELM	500	0.69	100	95.45

As observed in Table IV, compared with the other algorithms, FIG-ELM's training time was unattainable. It was easy to find that SVM was over-fitting and had a long training time. The ELM and FIG-ELM algorithms' training accuracy and testing accuracy increased with the hidden nodes' increase, but ELM's training time was clearly longer than FIG-ELM's, which were 2.04s and 0.69s when the number of hidden nodes was 500. Moreover, FIG-ELM's training and testing accuracy were the best. Hence, judging from the simulation results, the FIG-ELM has stronger learning efficiency and generalizability than ELM, especially when there are a large amount of data and need many hidden nodes.

V. CONCLUSIONS

This paper combined an extreme learning machine (ELM) with fuzzy information granulation (FIG). The new approach not only improved the speed of a basic ELM algorithm that contained many hidden nodes, but also overcame the weakness of a basic ELM of low learning efficiency and generalizability. The effectiveness was demonstrated by simulating four examples: two regression problems (a regression dataset and a nonlinear function approximation) and two classification problems (breast cancer and character recognition). Experimental results show that the FIG-ELM algorithm has a higher potential of enhancing predictive accuracy and robustness and for reducing training time.

Further efforts will be made to achieve a higher efficient level. We also need to improve the ELM algorithm, such as online sequential ELMs, etc.

REFERENCES

[1] Huang G.B., Zhu Q.Y., Siew C.K., Extreme learning machine: Theory and applications. Neurocomputing (S0925-2312), 2006, 70(1): 489-501. http://dx.doi.org/10.1016/j.neucom.2005.12.126

[2] Huang G.B., Ding X.J., Zhou H.M., Optimization method based extreme learning machine for classification. Neurocomputing, 2010, 74(1-3): 155-163. http://dx.doi.org/10.1016/j.neucom.2010.02.019

[3] Huang G.B., Ding X.J., Zhou H.M., et al. Extreme learning machine for regression and multiclass classification. IEEE Trans. On Systems, Man, and Cybernetics, 2012, 42(2): 513-529. http://dx.doi.org/10.1109/TSMCB.2011.2168604

[4] Huang G.B.,Zhu Q.Y., C.-K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of International Joint Conference on Neural Networks (IJCNN2004), Budapest, Hungary, 2004.

[5] Zhang R.D., Wang Q.S., Neural network based multi-step predictive control for nonlinear system. Control and Decision, 2005, 20(3): 332-336.

[6] Cui G.M., Sun T., Zhang Y., Application of Support Vector Machine (SVM) in Prediction of Molten Iron Temperature in Blast Furnace. Control Engineering of China, 2013, 20(5): 809-817.

[7] Huang G.B., Siew C.K., Extreme Learning Machine with Randomly Assigned RBF Kernels. International Journal of Information Technology, 2005.

[8] Liao R.J., Zheng H.B., Grzybowski S. et al, Fuzzy information granulated particle swarm optimisation-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers. IET Electric Power Applications, 2011, 5(2): 230-237. http://dx.doi.org/10.1049/iet-epa.2010.0103

[9] Ding S.F., Han Y.Z., Yu J.Z. et al, A fast fuzzy support vector machine based on information granulation. Neural Comput & Applic, 2013, 23(suppl 1): S139-S144. http://dx.doi.org/10.1007/s00521-012-1276-8

[10] Han F., Yao H.F., Ling Q.H., An improved extreme learning machine based on particle swarm optimization. Neurocomputing, 2011, 116: 87-93. http://dx.doi.org/10.1016/j.neucom.2011.12.062

AUTHORS

Xiafu Lv is with the Key Laboratory of Industrial Internet of Things & Network Control, MOE, Chongqing University of Posts and Telecom, Chongqing CO 400065 China (e-mail: Lvxf@ cqupt.edu.cn).

Jun-peng CHEN is with the Key Laboratory of Industrial Internet of Things & Network Control, MOE, Chongqing University of Posts and Telecom, Chongqing CO 400065 China (e-mail: 1328222765@qq.com).

Lei LIU is with the Key Laboratory of Industrial Internet of Things & Network Control, MOE, Chongqing University of Posts and Telecom, Chongqing CO 400065 China (e-mail: 115099392@ qq.com).

Bo-hua WANG is with the Key Laboratory of Industrial Internet of Things & Network Control, MOE, Chongqing University of Posts and Telecom, Chongqing CO 400065 China (e-mail: 770623387@ qq.com).

Yong WANG is with the College of Mobile Telecommunications, Chongqing University of Posts and Telecom Chongqing CO 401520 China (e-mail: 1287070946@ qq.com).

Submitted 21 July 2015. Published as rsubmitted by the authors 10 October 2015.