

A Novel Extreme Learning Machine Based on Hybrid Kernel Function

Shifei Ding^{1,2}, Yanan Zhang¹, Xinzheng Xu¹, Lina Bao¹

1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, 221116 China
Email: zhangyanan.07@163.com

2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190 China
Email: {dingsf, xxzheng}@cumt.edu.cn

Abstract—Extreme learning machine is a new learning algorithm for the single hidden layer feedforward neural networks (SLFNs). ELM has been widely used in various fields and applications to overcome the slow training speed and over-fitting problems of the conventional neural network learning algorithms. ELM algorithm is based on the empirical risk minimization, without considering the structural risk and this may lead to over-fitting problems and at the same time, it is with poor controllability and robustness. For these deficiencies, an optimization method is proposed in this paper, a novel extreme learning machine based on hybrid kernel function (HKELM). The method constructs a hybrid kernel function with better performance by fully combining local kernel function strong learning ability and global kernel function strong generalization ability. Compared with traditional ELM, the results show that this method can effectively improve the ELM classification results, avoid local minimum, with better generalization, robustness, controllability and faster learning rate.

Index Terms—Hybrid Kernel Function; Extreme Learning Machine; Global Kernel Function; Local Kernel Function

I. INTRODUCTION

Artificial Neural Network has already become an indispensable learning method in the field of artificial intelligence. Neural network is with parallel and distributed information processing network architecture, and has a strong nonlinear mapping capability and adaptive, self-learning, robustness and fault-tolerance characteristics. Nowadays, neural networks have been widely applied to pattern recognition, image processing, intelligent expert system for control and other area.

Extreme Learning Machine is a single hidden layer feedforward neural network fast learning algorithm proposed in 2006 by Huang et al [1]. The algorithm is mainly for classification and regression. The thought is during the network training, first through continuous testing to set the number of hidden nodes, the random assignment of input weights and hidden layer deviation, obtained the output layer right value by the least squares method. To complete whole learning process only need one mathematical change without iterative. The training

speed compared to traditional BP algorithm based on gradient descent has been significantly improved [2].

Currently, ELM has been applied in many areas, such as, E. Romero, et al [3] proposed an error minimization speed learning machine (EM-ELM) to establish a way to add hidden layer nodes in random order to construct SLFNs easily. This can update the incremental minimum output weight training error square sum. In statistics, error minimization ELM significantly improves generalization performance. M. R. Daliri [4] used the combination of genetic algorithms and fuzzy ELM hybrid system to clinical practice. Y.Q. Chang, et al [5] proposed an improved ELM algorithm to deal with the disadvantage of ELM method excessive number of hidden layer neurons problems. The method added classification neurons to the hidden layer of SLFN. W. W. Zong, et al [6] proposed a weighted extreme learning machine for imbalance learning to deal with data with imbalanced class distribution. However, at present the ELM algorithm still have some drawbacks, the theoretic basis of the algorithm also need deeper mining, and at the same time, selection and optimization method of ELM kernel function needs to be further studied.

Similar with the traditional neural network, the selection of ELM kernel function is also a core issue. Selected the appropriate kernel function largely affect the performance of ELM, including the generalization performance and learning performance. When conducting training the neural network, the selection of the kernel function greatly impact to the network learning rate and learning effects. Aiming at the above problems in ELM, this paper proposes a kind of extreme learning machine learning method based on hybrid kernel function, called Hybrid Kernel Extreme Learning Machine, HKELM for short. HKELM method first will combine the local kernel function which has a better learning ability with the global nuclear function which has better generalization ability, form a hybrid kernel function. Then use the hybrid kernel as a kernel of ELM for networking training. Hybrid kernel combines the advantages of both kernel and has a good learning ability and strong generalization ability, improves the efficiency and accuracy of ELM.

II. ELM THEORY

A. Model of ELM

For N arbitrary distinct samples (x_i, t_i) , where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, what's more $(x_i, t_i) \in R^n \times R^m (i = 1, 2, \dots, N)$, standard SLFNs with \aleph hidden nodes and activation function $f(x)$ are mathematically models as

$$\sum_{i=1}^{\aleph} \beta_i f_i(x_j) = \sum_{i=1}^{\aleph} \beta_i f(a_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \tag{1}$$

Where $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$ is the weight vector connecting the i th hidden node and the input nodes, and b_i is the threshold of the i th hidden node. $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector connecting the i th hidden node and the output nodes. $a_i \cdot x_j$ represents the inner product of a_i and x_j , and the activation function usually choose "Sigmoid", "Sine", "RBF".

The above equation (1) can be written compactly as $H\beta = T$ (2)

where $H(a_1, \dots, a_{\aleph}, b_1, \dots, b_{\aleph}, x_1, \dots, x_N)$

$$H = \begin{bmatrix} f(a_1 \cdot x_1 + b_1) & \dots & f(a_{\aleph} \cdot x_1 + b_{\aleph}) \\ \dots & \dots & \dots \\ f(a_1 \cdot x_N + b_1) & \dots & f(a_{\aleph} \cdot x_N + b_{\aleph}) \end{bmatrix}_{N \times \aleph}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \dots \\ \beta_{\aleph}^T \end{bmatrix}_{\aleph \times m}, T = \begin{bmatrix} t_1^T \\ \dots \\ t_N^T \end{bmatrix}_{N \times m}$$

H is called the hidden layer output matrix of the neural network; the i th column of H is the i th hidden node output with respect to inputs x_1, x_2, \dots, x_N .

Theorem proving [7], as long as the number of hidden nodes is enough, when the activation function $f(x)$ is infinitely differentiable at any interval the parameters of the network does not all need to adjust. When the training starts, SLFN randomly assigns to the input connection weights a and hidden layer node bias b , moreover, while the training process unchanged it can approximate any continuous function. Generally, in order to get good generalization performance, take $\aleph = N$.

When the input weights and hidden layer bias are determined in accordance with the random assignment, according to the input samples can get the hidden layer output matrix H . Therefore, training SFLN is converted into solving linear equations $H\beta = T$ least squares solution.

$$\begin{aligned} & \|H(a_1, \dots, a_{\aleph}, b_1, \dots, b_{\aleph})\hat{\beta} - T\| \\ & = \min_{\beta} \|H(a_1, \dots, a_{\aleph}, b_1, \dots, b_{\aleph})\beta - T\| \end{aligned} \tag{3}$$

The above equation (3) least squares solution of the above liner system is

$$\hat{\beta} = H^\dagger T \tag{4}$$

In the equation (4), H^\dagger represents Moore-Penrose [8] generalized inverse of the hidden layer output matrix H . Usually, the optimal solution $\hat{\beta}$ contains the following features:

- (1) According to $\hat{\beta}$, the algorithm can gain the minimal training error;
- (2) Can get the optimal generalization capability of the minimum paradigm of the output connection weights and network;
- (3) $\hat{\beta}$ is unique. This can avoid producing the local optimal solution.

ELM through selecting the appropriate kernel function, classification of nonlinear mapping in the original space into a high-dimensional space, achieved in high-dimensional spaces of linear classification.

B. Kernel Function

According to the theory of pattern recognition, low-dimensional space linear inseparable mode by nonlinear mapping to high-dimensional feature space can achieve linear separable. However, if we directly use this technology in high-dimensional space for classification or regression, then there will be issues on determining the form and parameters of the nonlinear mapping functions and feature space dimension. What's more, in the high-dimensional feature space operation the biggest obstacle is the existence of "dimensions of disaster". The kernel function technology can effectively solve these problems.

As early as 1964, Aizermann, in the study on the potential function method on the technology introduced kernel function theory to the field of machine learning and it is ripe for the application of neural network algorithm model. Kernel function method is using arbitrary random vector X in the n dimensional vector space mapped to a high dimensional feature space $F : x \rightarrow \Phi(x) \in F$ with a nonlinear transformation and can achieve a high dimensional feature space linear classification. In high dimensional feature space F , the interaction between each coordinate component is limited to the inner product linear learning algorithm, does not require specific forms of nonlinear transformation, as long as the kernel function to replace the inner product in the linear algorithm satisfy the Mercer condition, you can get the original input space corresponding nonlinear algorithm.

More of the kernel function in the present study there are 3 main categories:

- 1) Polynomial kernel function $K(x, x_i) = (\gamma(x \cdot x_i) + r)^d, \gamma > 0$ (5)
- 2) Perceptron kernel function

$$K(x, x_i) = \tanh(v(x \cdot x_i) + c) \tag{6}$$

3) Gauss RBF nuclear function

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{7}$$

4) Sigmoid function

$$K(x) = \frac{1}{1 + \exp(-x)} \tag{8}$$

In formula (5) to (7), d, c, σ are real constant parameters.

C. Hybrid Kernel Function

There are many type of kernel function and the commonly used kernel function summed up mainly can be divided into two main types of local kernel function and global nuclear function. For example, RBF function is a typical local kernel function, the Perceptron kernel function and polynomial kernel function are two typical global nuclear function. Local kernel functions only in test point near the region class has an impact on the data points, and global nuclear function allows the kernels away from the test input data point values also affect [9]. Because the local kernel has high learning ability but the generalization performance weak, on the contrary, the global nuclear function has strong generalization performance but weak learning ability. Therefore, these two types of nuclear function mixed together to form a hybrid kernel function.

The traditional method of constructing kernel functions is based on the idea of Mercer's theorem which was proposed in 1909 by R. Wille [10].

Theorem. If the function K is mapped on $R^n \times R^n \rightarrow R$ (mapped to the real number field from two n dimensional vector). Then if K is a Mercer kernel, if and only if for the training samples $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, the corresponding kernel matrix is symmetric positive semidefinite.

Simple kernel function constructed complex hybrid kernel function, namely the hybrid kernel function still satisfy the Mercer theorem of the kernel function.

Set K_1, K_2 is defined in the kernel function on $X \times X$, f is real-valued functions on $X: \Phi(x): X \rightarrow R^N$, K_3 is a kernel function on $R^N \times R^N$, $a \in R^+$, B is a $n \times n$ dimensional positive semidefinite symmetric matrix, then by Mercer theorem, the functions as follows are the kernel function:

- 1) $K(x, z) = K_1(x, z) + K_2(x, z)$
- 2) $K(x, z) = f(x)f(z)$
- 3) $K(x, z) = aK_1(x, z)$
- 4) $K(x, z) = K_1(x, z)K_2(x, z)$
- 5) $K(x, z) = K_3(\Phi(x), \Phi(z))$
- 6) $K(x, z) = x^T Bz$

Proof. Set matrix K_1, K_2 defined on a finite set of points $\{x_1, \dots, x_n\}$, for any vector $\alpha \in R^n$, K is a positive semidefinite matrix and the necessary and sufficient

condition is all α must satisfy $\alpha^T K \alpha \geq 0$. It can be concluded that $\alpha^T (K_1 + K_2) \alpha = \alpha^T K_1 \alpha + \alpha^T K_2 \alpha \geq 0$, then $K_1 + K_2$ is positive semidefinite, that is $K_1 + K_2$ meet Mercer theorem, so it is kernel function, namely, $K(x, z) = K_1(x, z) + K_2(x, z)$ is kernel function. Other similarly certification, does not enumerate here.

Certified by the above we can see that the function $K = \lambda K_1 + (1 - \lambda) K_2$, $\lambda \in (0, 1)$ is a mixed function satisfied the Mercer condition. In this paper, we combine the RBF kernel function

$K_1(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right)$ and the Perceptron kernel

function $K_2(x, x_i) = \tanh(v(x \cdot x_i) + c)$ to form a hybrid kernel function. RBF kernel function is a typical local kernel function and the Perceptron kernel function is also a commonly used global nuclear function. In the paper, we do a liner combination with the two kernel function to form a hybrid kernel function, thus it can well integrate superiority of both kernels. Then it can improve both the learning ability of the kernel function and its generalization ability. Apply it to the ELM thus achieve good classification performance.

Radial basis function (RBF) with great learning capacity because it only has impact on test point near regional data. As shown in Figure 1, when the function parameter σ takes different values, the RBF kernel function learning results in the test points 0.05.

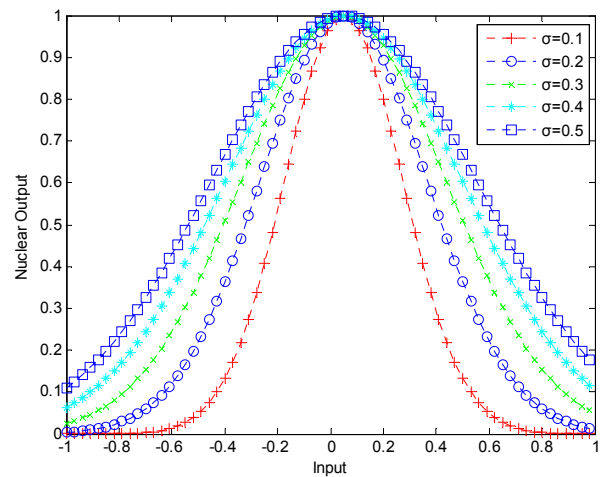


Figure1 Curves of RBF kernel function in the test point 0.05

It can be seen through the figure, RBF kernel has a good learning ability only for the vicinity of the test point 0.05, but relatively poor generalization ability. From the diagram we can also see that the smaller the value of σ , the RBF kernel function will get the stronger learning capacity, while the generalization ability will be correspondingly weakened. Generally, we make $\sigma \in [0.1, 1]$.

And the Perceptron kernel function is global nuclear function with strong generalization capability. As shown in Figure 2, when selecting different parameters, the

learning effect of the Perceptron kernel function in the test point.

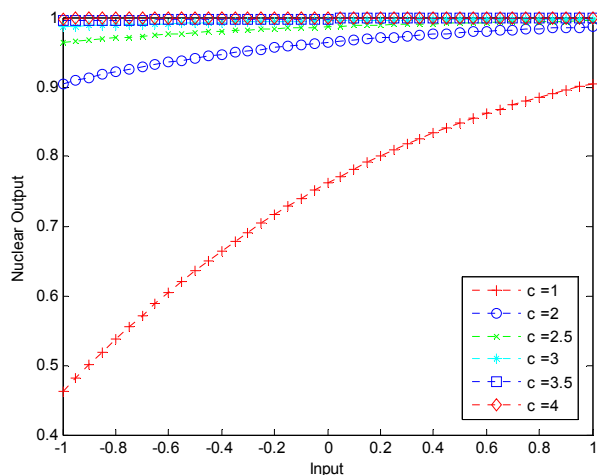


Figure2 Curves of Perceptron kernel function in the test point 0.05

In the figure above, $v = 10$, when c respectively different values, as can be seen from the diagram, on the condition that the selection of the parameters are appropriate, Perceptron kernel function has a strong generalization ability and able to has a role at the test points near and far for the entire range of data. However, on the contrary, the learning ability in the test point 0.05 is not very obvious. This indicates that the learning ability of the Perceptron kernel function is weak. After several experiments found that to a certain extent, the value of v affects the Perceptron kernel function generalization effect. When $5 \leq v \leq 15$ the results are more appropriate, therefore, in this article for convenience to illustrate sets $v = 10$, furthermore, when $v = 10$, as shown above, on the condition that $c \geq 3$, the nuclear output value able to achieve stability will not produce changes.

With the above can know that the Perceptron kernel function generalization capacity superior to the RBF kernel function, but poorer learning ability. Thus, combining these two kernel functions to form a hybrid kernel function which with a good learning a good learning ability and generalization ability. At the same time to ensure that the hybrid kernel does not change the rationality in the original mapping space and then the proportion coefficients sum of two kernel functions is 1. Based on this idea, this paper presents a hybrid kernel function.

$$K(x, x_i) = \lambda \exp\left(\frac{\|x - x_i\|^2}{\sigma^2}\right) + (1 - \lambda) \tanh(v(x \cdot x_i) + c) \tag{9}$$

Then we can learn that the formula (9) is a kernel function through the theorems and proofs, where λ is the coefficient proportion of RBF kernel function in the hybrid kernel function and it can be referred to as the mixing coefficients. Figure 3 is a hybrid kernel function when the parameters $v = 10$, $c = 3$, $\sigma = 0.5$, λ take different values on the test point the output curve graph.

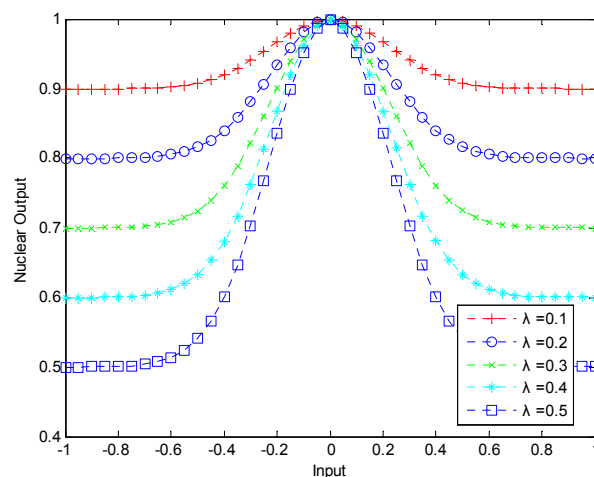


Figure3 Curves of hybrid kernel function in the test point 0.05

III. ELM BASED ON HYBRID KERNEL FUNCTION (HKELM)

A. HKELM Algorithm Learning Steps

In the paper, we build an extreme learning machine model based on hybrid kernel function called HKELM for short. In the HKELM model, the hybrid kernel function combines strong learning ability of local kernel function and strong generalization ability of global nuclear function, with better learning performance. Specific steps of the algorithm are as follows:

Step 1: Define network. Define the number of hidden layer nodes and randomly assign to input weights and implied bias.

Step 2: Build Model. Introduce the hybrid kernel function to build ELM learning model.

Step 3: Samples training. Train the sample data sets.

Step 4: Performance testing. Use the testing datasets to test HKELM generalization performance and learning performance.

B. HKELM Algorithmic Performance analysis

The HKELM model presented in this article, mainly do the ELM kernel function optimization. By combining local kernel function and global nuclear function to form the hybrid kernel function and then placed it in the ELM for training and learning. The performance of the algorithm is mainly from the following two aspects to analyze.

1) Analysis of the learning ability of the algorithm

Compared with the traditional extreme learning algorithm, learning models in this article imports hybrid kernel function and combines with RBF strong learning ability. What's more, the ELM operation process only needs a mathematical transformation to complete. This significantly improves the learning efficiency and accuracy.

2) Analysis of the generalization ability of the algorithm

A good learning model should have a good generalization performance. Traditional ELM is based on the empirical risk minimization theory and the hybrid

kernel function combined with the global nuclear function played a significant role in learning model generalization performance. Therefore this learning model has better generalization performance.

IV. EXPERIMENTAL ANALYSIS

In order to verify the performance of HKELM model, in this paper, in Matlab environment we do experiment on several classification datasets of UCI database and compare the results to verify the learning effect when ELM select RBF kernel function, Perceptron kernel function and hybrid kernel function as a kernel. The data sets used in the experiments mainly includes: btsc (Blood Transfusion Service Center) dataset, wine dataset and HeartStatlog dataset. The basic characteristics are shown in Table I.

TABLE I
BASIC CHARACTERISTICS OF THE THREE DATASETS

| Data sets | Class | Samples | Condition attributes |
|--------------|-------|---------|----------------------|
| btsc | 2 | 748 | 4 |
| wine | 3 | 178 | 13 |
| HeartStatlog | 2 | 270 | 13 |

In order to reduce the impact of human factors on the experimental results, the selected sample data sets were randomly divided into two equal parts, respectively as the training datasets and testing datasets. First of all, the training datasets as input respectively do the training processes to choose these three different kernel functions of ELM. And then test the trained network with testing datasets. The following table shows when ELM respectively selects Perceptron kernel, RBF kernel and hybrid kernel, the comparison of the datasets classification results. Each selected parameters of the kernel function are the optimal value chosen in many times experiments, when the mixing coefficient $\lambda = 0.1$ the hybrid kernel function influence on ELM classification results.

TABLE II
COMPARISON OF THREE KINDS OF KERNEL KUNCTIONS ELM CLASSIFICATION RESULTS

| Kernel function | Classification accuracy (%) | | |
|-----------------|-----------------------------|-------|--------------|
| | btsc | wine | HeartStatlog |
| Perceptron-ELM | 70.86 | 79.78 | 68.89 |
| RBF-ELM | 62.03 | 85.39 | 46.67 |
| HKELM | 78.61 | 97.75 | 83.70 |

As can be seen from the above table, for btsc dataset and HeartStatlog dataset, the Perceptron kernel function showed better classification results than the RBF kernel function. This reflects as global nuclear function, perceptron ELM learning model has strong generalization ability. But for wine dataset, the RBF kernel function classification effect is better than the Perceptron kernel function. This shows that as a local kernel function, the RBF kernel function ELM learning model has strong learning ability. However, in General, when ELM selects the mixed function of the two as the kernel function, the hybrid kernel ELM learning model whether generalization performance or learning performance are much higher than the former two. This confirms that the

HKELM has good learning performance and generalization performance.

Then Following further discuss effects of different values of mixing coefficient λ on the performance of the hybrid kernel function. In the hybrid kernel function, mixing coefficient λ is to adjust the proportion of two kinds kernel function. Theorem and proof show that for the hybrid kernel function

$$K(x, x_i) = \lambda \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) + (1 - \lambda) \tanh(v(x \cdot x_i) + c) ,$$

$\lambda \in (0, 1)$ when λ is small the RBF kernel function occupy a high proportion of the hybrid kernel function. When λ is large the Perceptron kernel functions occupy a high proportion of the hybrid kernel function. By transforming the value of λ to study the impact of λ on generalization ability and learning ability of HKELM. The results are shown in Table III.

TABLE III
EFFECT OF MIXING COEFFICIENT λ TO HKELM CLASSIFICATION RESULTS

| Value of λ | Classification accuracy (%) | | |
|--------------------|-----------------------------|-------|--------------|
| | btsc | wine | HeartStatlog |
| 0.1 | 78.61 | 97.75 | 83.70 |
| 0.2 | 77.54 | 96.63 | 82.96 |
| 0.3 | 76.74 | 95.51 | 79.26 |
| 0.4 | 76.20 | 94.38 | 80.74 |
| 0.5 | 75.94 | 94.38 | 76.30 |
| 0.6 | 76.74 | 93.26 | 77.78 |
| 0.7 | 76.47 | 91.01 | 81.48 |
| 0.8 | 77.81 | 92.13 | 80.00 |
| 0.9 | 77.27 | 95.51 | 82.22 |

As can be seen from the above table, when λ is small the RBF kernel function occupy a high proportion of the hybrid kernel function and HKELM has strong learning ability. With the increase of λ , the proportion of RBF kernel function is gradually reduced and the Perceptron kernel functions occupy a high proportion, HKELM generalization performances gradually increase. On the whole, combined with the experimental data can be informed that the learning performance of HKELM is better than performance of the single the Perceptron-ELM and the RBF-ELM. Experimental results show that HKELM guarantee both the good learning ability and good generalization performance. In this paper, the value of λ is the average of many times experiment.

In order to be more intuitive and convenient comparison, respectively select test results of the three different data sets Figure on Perceptron-ELM, RBF-ELM and HKELM.

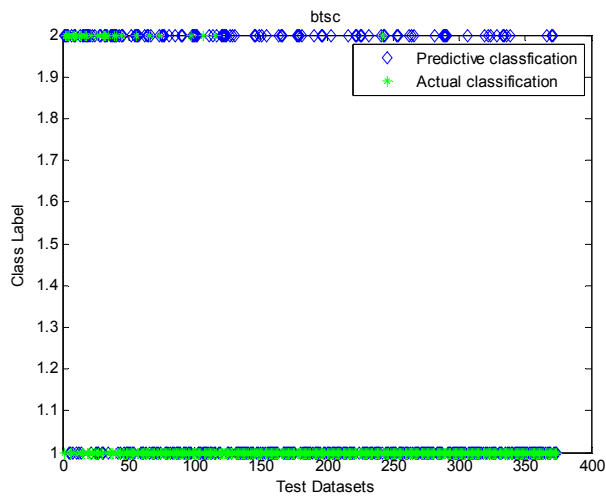


Figure4 Perceptron-ELM test result

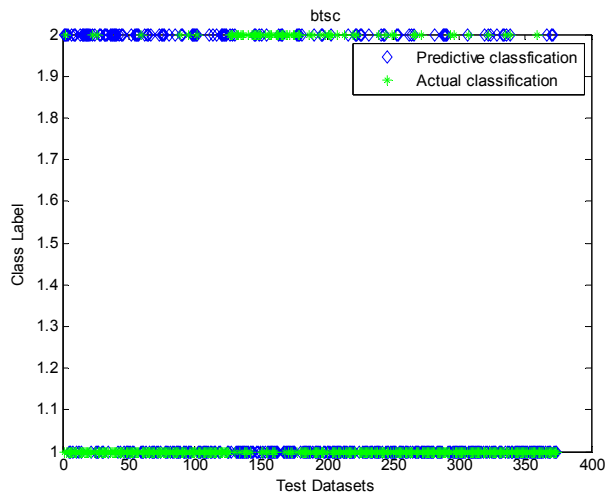


Figure5 RBF-ELM test result

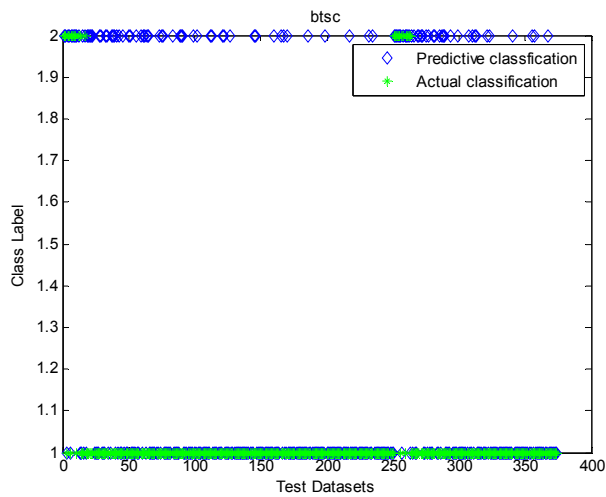


Figure6 HKELM test result

Figure 4 ~ Figure 6 are respectively the RBF-ELM, Perceptron-ELM and HKELM processing with btscc dataset. The actual classification and prediction of the testing dataset classification results comparison chart. Can be seen from the figure, the classification accuracy of HKELM is higher than the RBF-ELM and Perceptron-ELM.

Following here are the RBF-ELM, Perceptron-ELM and HKELM in the processing of wine and HeartStatlog datasets classification results figures.

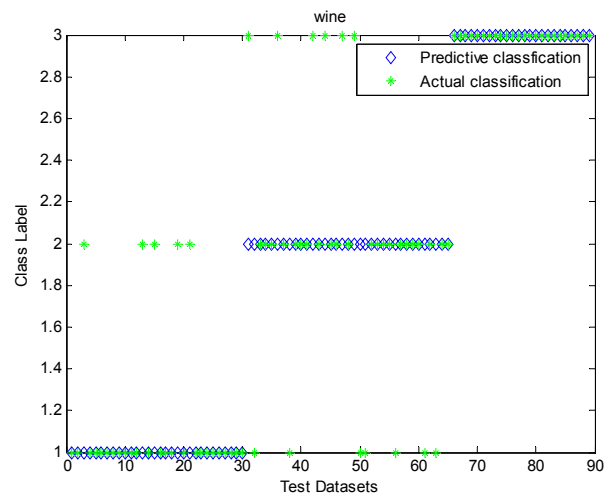


Figure7 Perceptron-ELM test result

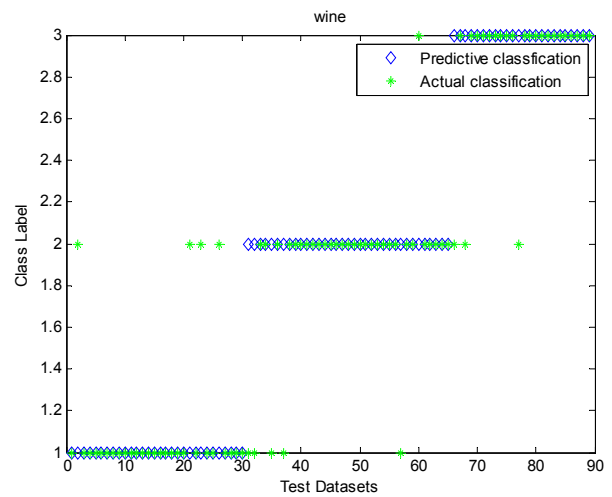


Figure8 RBF-ELM test result

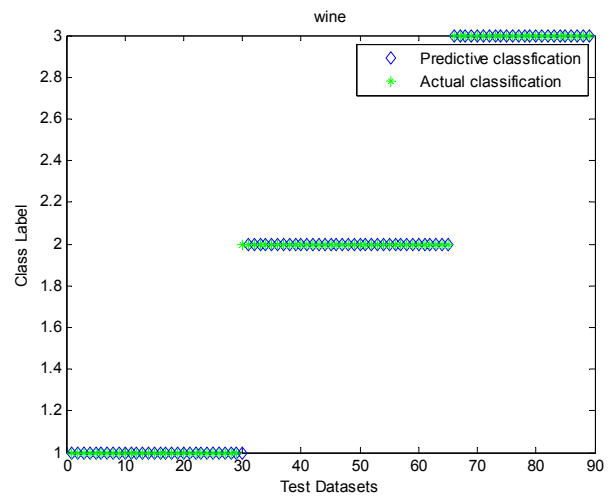


Figure9 HKELM test result

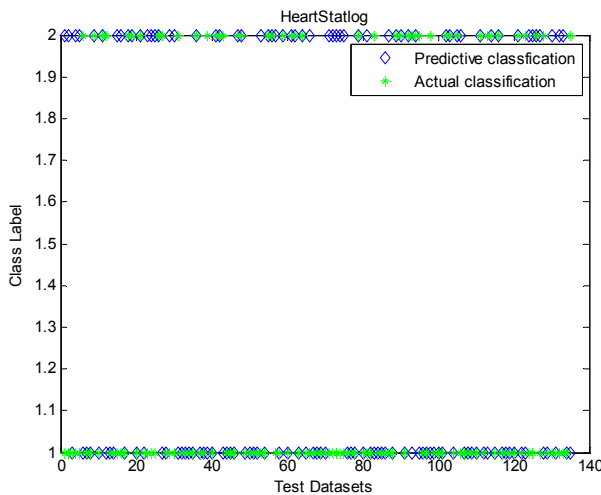


Figure10 Perceptron-ELM test result

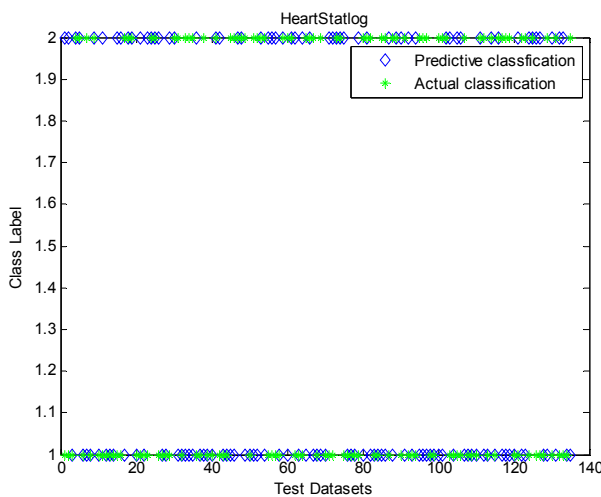


Figure11 RBF-ELM test result

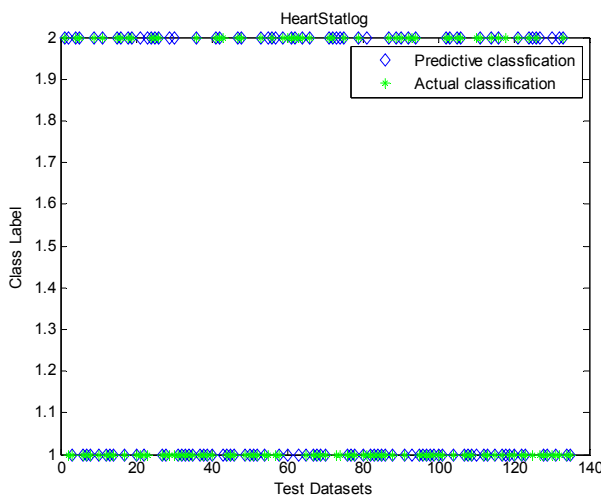


Figure12 HKELM test result

It can be seen from the figure above, HKELM compared to traditional ELM has better generalization performance and learning performance.

V. CONCLUSIONS

In order to improve the performance of ELM, a hybrid extreme learning machine was proposed in this paper.

Through a combination of the common global nuclear function and local kernel function to construct a hybrid kernel functions combines the advantages of the two nuclear functions. Both high learning performance and a good generalization performance, and through experiments confirmed the feasibility of the HKELM. Selection of local kernel function and global nuclear function can also consider other kernel function mixed to constitute a mix of different kernel functions for classification learning. However, the mixed kernel function parameter selection, mixing coefficient values of further research is needed and it is also the next step of the work.

ACKNOWLEDGMENT

This work is supported by the National Key Basic Research Program of China (No. 2013CB329502), the National Natural Science Foundation of China (No.41074003), the Opening Foundation of the Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (IIP2010-1), and The Opening Foundation of Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications.

REFERENCES

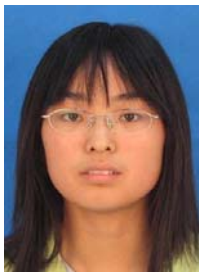
- [1] G.B. Huang, Q.Y. Zhu, C.K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006, 70 (1-3): 489-501.
- [2] W.Y. Deng, Q.H. Zheng, S.G. Lian et al. Ordinal extreme learning machine. *Neurocomputing*, 2010, 74 (1-3): 447-456.
- [3] E. Romero, R. Alquezar. Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks. *Neural Networks*, 2012, 25,122-129.
- [4] M.R. Daliri. A Hybrid Automatic System for the Diagnosis of Lung Cancer Based on Genetic Algorithm and Fuzzy Extreme Learning Machines. *Journal of Medical System*, 2012, 36 (2): 1001-1005.
- [5] Y. Q. Chang, Y. C. Li, F. L. Wang et al. Soft Sensing Modeling Based on Extreme Learning Machine for Biochemical Processes. *Journal of System Simulation*, 2007, 19(23):5587-5590.
- [6] W. W. Zong, G. B. Huang, Y. Q. Chen. Weighted extreme learning machine for imbalance learning. *Neurocomputing*, 2013, 101: 229-242.
- [7] G.B. Huang. Learning capability and storage capacity of two-hidden-layer feedforward network. *IEEE Transactionson Neural Networks*, 2003, 14(2):274-281.
- [8] N.Y. Liang, G.B. Huang. A fast and accurate online sequential learning algorithm for feed forward networks. *IEEE Transactionson NeuralNetworks*, 2006, 17(6):1411-1423.
- [9] Inderjit S. Dhillon, Dharmendra S. Modha. Concept decomposition for large sparse text data using clustering. *Machine Learning*, 2001, 42(1):143-175.
- [10] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival I ed. *Ordered Sets*. Reidel, Dordrecht, 1982. 445-470.



Shifei Ding is a professor and Ph.D. supervisor at China University of Mining and Technology. His research interests include intelligent information processing, pattern recognition, machine learning, data mining, and granular computing et al. He has published 3 books, and more than 100 research papers in journals and international conferences. He received

his B.Sc., M.Sc. degree in mathematics and computer science in 1987, 1998 respectively from Qufu Normal University, received Ph.D. degree in computer science from Shandong University of Science and Technology in 2004, and received post Ph.D. degree in computer science from Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, and Chinese Academy of Sciences in 2006.

In addition, prof. Ding is a senior member of China Computer Federation (CCF), and China Association for Artificial Intelligence (CAAI). He is a member of professional committee of Artificial Intelligence and Pattern Recognition, CCF, professional committee of distributed intelligence and knowledge engineering, CAAI, professional committee of machine learning, CAAI, and professional committee of rough set and soft computing, CAAI. He is an associate Editor-in-Chief for International Journal of Digital Content Technology and its Applications (JDCTA), acts as an editor for Journal of Convergence Information Technology (JCIT), International Journal of Digital Content Technology and its Applications (JDCTA) et al.



Zhang yan is currently a graduate student now studying in China University of Mining and Technology, School of Computer Science and Technology, Xu Zhou, Jiang Su, China.

She received her bachelor's degree from China University of Mining and Technology in 2011. Her research interests include intelligent information processing, pattern recognition, machine learning, and granular computing et al.



Xu Xinzheng received his BS degree in electrical engineering from Shandong University of Science and Technology in 2002 and received the M.D. and Ph.D. degrees in Computer Application Technology from Xiamen University in 2005 and China University of Mining and Technology in 2012, respectively.

He is currently an associate professor at School of Computer Science and Technology, China University of Mining and Technology. Nowadays, he is a member of China Computer Federation, and China Association for Artificial Intelligence. His research interests include intelligent information processing, pattern recognition, machine learning, and granular computing et al.