

Research Article

Enhancement of ELM by Clustering Discrimination Manifold Regularization and Multiobjective FOA for Semisupervised Classification

Qing Ye,¹ Hao Pan,¹ and Changhua Liu²

¹School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430000, China

²Yangtze University College of Technology and Engineering, Jingzhou 430023, China

Correspondence should be addressed to Changhua Liu; l2669963@qq.com

Received 8 December 2014; Revised 8 May 2015; Accepted 12 May 2015

Academic Editor: Cheng-Jian Lin

Copyright © 2015 Qing Ye et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel semisupervised extreme learning machine (ELM) with clustering discrimination manifold regularization (CDMR) framework named CDMR-ELM is proposed for semisupervised classification. By using unsupervised fuzzy clustering method, CDMR framework integrates clustering discrimination of both labeled and unlabeled data with twinning constraints regularization. Aiming at further improving the classification accuracy and efficiency, a new multiobjective fruit fly optimization algorithm (MOFOA) is developed to optimize crucial parameters of CDME-ELM. The proposed MOFOA is implemented with two objectives: simultaneously minimizing the number of hidden nodes and mean square error (MSE). The results of experiments on actual datasets show that the proposed semisupervised classifier can obtain better accuracy and efficiency with relatively few hidden nodes compared with other state-of-the-art classifiers.

1. Introduction

Recently, ELM [1, 2] shows better performance than traditional gradient-based learning methods and support vector machine (SVM) [3, 4] in regression and classification applications due to its faster learning capacity. As a supervised learning algorithm, the applicability of ELM is seriously restrained [5]. In actual applications, unlabeled data are easy to obtain while the acquisition of labeled data is time consuming and hard. Based on this, it is imperative to extend ELM to achieve semisupervised classification.

Manifold regularization is a frequently used semisupervised learning method based on smoothness assumption [6]. LapRLS [7] and LapSVM [8, 9] based on manifold assumption are frequently used semisupervised learning algorithm. However, manifold regularization is prone to sinking into misclassification in boundary area between several clusters because boundary instances in manifold structure are likely to belong to different classes [10]. Wu et al. [11] proposed semisupervised discrimination regularization (SSDR) for solving misclassification by utilizing discrimination of

labeled data in learning. However, due to the scarcity of labeled data, the improvement of misclassification is limited. Wang et al. [12] proposed discrimination-aware manifold regularization (DAMR) in which discrimination of the whole data is considered to improve accuracy. Yet, DAMR merely adopted binary cluster labels which are insufficient for multiclass problem. In view of this, an improved MR framework named clustering discrimination manifold regularization (CDMR) which integrates clustering discrimination of both labeled and unlabeled data with twinning constraints regularization is proposed, and a semisupervised ELM with CDMR framework termed CDMR-ELM is finally developed. The proposed novel framework can effectively avoid boundary misclassification which frequently occurred in manifold regularization and improve the classification accuracy by combining the clustering discrimination with twinning constraints regularization containing lower intra-cluster compactness and higher intercluster separability.

FOA is a global optimization searching method based on food finding behavior of fruit fly with the advantages of simplicity and being easy to understand [13, 14]. This

paper develops an improved variant of FOA named multiobjective fruit fly optimization algorithm (MOFOA) to optimize crucial parameters of CDMR-ELM consisting of the number of hidden nodes and trade-off parameters for further improving the classification accuracy and efficiency. The MOFOA employs MSE to evaluate fitness function and adopts adaptively reduced search area for decision variable to alleviate the possibility of sinking into local extremum and prematurity [15, 16]. Above all, unlike traditional FOA-ELM which implements optimization iteration with fixed number of hidden nodes, MOFOA is based on two objectives: simultaneously minimizing the number of hidden nodes and MSE which can obtain a set of optimal parameters to increase classification accuracy with fewer hidden nodes to reduce computational complexity and enhance efficiency.

The rest of this paper is organized as follows: Section 2 introduces related basic theory. Section 3 proposes a novel CDMR framework and integrates it with ELM. Section 4 presents MOFOA to optimize the parameters of CDMR-ELM. Experimental setup and comparison results are given in Section 5. Section 6 is the conclusion of this paper.

2. Related Basic Theory

2.1. Extreme Learning Machine (ELM). Consider a training set containing N arbitrary distinct samples $\{(x_i, y_i)\}_{i=1, \dots, N}$. Here $x_i \in R^d$ and $y_i \in R^m$, where d and m represent the dimensions of input and output vector. The output of ELM with respect to sample x_i is determined as follows [17]:

$$f(x_i) = \sum_{j=1}^L \beta_j G(a_j, b_j, x_i), \quad i = 1, \dots, N, \quad (1)$$

where L is the number of hidden nodes, $G(\cdot)$ is hidden layer output function, and β_j is output weight connecting the j th hidden node to output layer. Input weight a_j and bias b_j of hidden nodes are randomly assigned in advance. Equation (1) can be converted into a compact form as follows:

$$H\beta = Y, \quad (2)$$

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L}, \quad (3)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \quad (4)$$

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m}$$

where H is output matrix of hidden layer and, by minimizing the square loss of predicted error and norm of weight, ELM analyzes the optimal output weight β as follows:

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \sum_{i=1}^N \|h(x_i) \beta - y_i\|^2. \quad (5)$$

2.2. Manifold Regularization Framework. Manifold regularization framework is built on manifold assumption that close points in the intrinsic geometry of marginal distribution P_x should share similar labels and can effectively solve problem of training dataset consisting of both labeled and unlabeled data [18]. Labeled data $\{(x_i, y_i)\}_{i=1, \dots, l}$ are generated according to probability distribution P and unlabeled data $\{x_j\}_{j=1, \dots, u}$ are drawn according to P_x of P . By minimizing the following cost function, manifold regularization framework can obtain an optimal classification function $f(\cdot)$:

$$\min_f \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f(x_i)) + \gamma_A \|f\|_k^2 + \gamma_I \|f\|_I^2, \quad (6)$$

where $V(\cdot)$ represents loss function and regularization term $\|f\|_k^2$ represents the complexity of classifier and regularization term $\|f\|_I^2$ which represents smoothness of sample distribution and it can be approximated as

$$\|f\|_I^2 = \frac{1}{2(u+l)^2} \sum_{i,j=1}^{l+u} w_{ij} \|f(x_i) - f(x_j)\|^2 \quad (7)$$

$$= \frac{1}{(u+l)^2} f^T L f,$$

where $1/(u+l)^2$ is normalization coefficient for the empirical estimate, $L = D - W$ is Laplacian matrix of the whole data, W is the weight matrix in which each element w_{ij} represents the similarity weight between $f(x_i)$ and $f(x_j)$, and D is a diagonal matrix in which $D_{ii} = \sum_{j=1}^{l+u} w_{ij}$.

2.3. Fruit Fly Optimization Algorithm (FOA). The steps of FOA are shown as follows.

Step 1. Randomly initialize the location of fruit fly: X_axis , Y_axis .

Step 2. Randomly generate the distance and direction for searching food by using osphresis of an individual: $X_i = X_axis + \text{RandomValue}$, $Y_i = Y_axis + \text{RandomValue}$.

Step 3. Estimate the distance D between each individual and origin and set the reciprocal of D as smell concentration judgment value S :

$$D(i) = \sqrt{X_i^2 + Y_i^2}, \quad (8)$$

$$S(i) = \frac{1}{D(i)}.$$

Step 4. Substitute $S(i)$ into smell concentration judgment function or fitness function of optimization to calculate the smell concentration $\text{Smell}(i)$ of individual fruit fly: $\text{Smell}(i) = F(S(i))$.

Step 5. Find out the individual fruit fly with maximal smell concentration: $[\text{bestSmell } \text{bestindex}] = \max(\text{Smell})$ in which bestindex is the location of best individual.

Step 6. Reserve best smell concentration value and corresponding coordinate of best individual: $F_{\text{best}} = \text{bestSmell}$, $X_{\text{axis}} = X(\text{bestindex})$, and $Y_{\text{axis}} = Y(\text{bestindex})$.

Step 7. Repeat Step 2 to Step 5 to execute iterative optimization until termination arrived and judge whether the smell concentration is better than previous one, if so, execute Step 5.

3. The Proposed Classifier: CDMR-ELM

3.1. CDMR Framework. In this paper, we consider a multi-class dataset with l labeled data $\{(x_i, y_i)\}_{i=1, \dots, l}$ and u unlabeled data $\{x_j\}_{j=l+1, \dots, l+u}$. Firstly, with the purpose of obtaining the clustering discrimination of the whole data, utilize unsupervised fuzzy clustering method [19] to divide the whole dataset into C fuzzy clusters which can effectively reflect the underlying cluster structure. Preserve all cluster labels to form a cluster vector with dimension of $(l + u)$ expressed as $T^C = [t_1^c, \dots, t_{l+u}^c]$, where t_j^c which is between 1 and C represents the fuzzy clustering label of the j th data. In order to fully consider the reliability of clustering result during the learning process, define a membership vector $M = [m_1^c, \dots, m_{l+u}^c]^T$ in which the element m_j^c represents the memberships degradation of the j th data defined as follows:

$$m_j^c = \begin{cases} 1, & \text{if } x_j \text{ is labeled data} \\ \mu, & \text{else,} \end{cases} \quad \mu \in [0, 1], \quad (9)$$

where μ is inversely proportional to distance between point and center of corresponding fuzzy cluster. Then, set $K = MM^T$ to describe the reliability of clustering. The clustering discrimination matrix S^c is defined on the basis of clustering labels and clustering reliability matrix K . The element S_{ij}^c of clustering discrimination matrix S^c represents whether i th instance and j th instance belong to the same fuzzy clustering and is defined as follows:

$$S_{ij}^c = \begin{cases} 1, & \text{if } t_i^c = t_j^c \\ 1, & \text{if } t_i^c \neq t_j^c, K_{ij} \leq 0.5 \\ -1, & \text{if } t_i^c \neq t_j^c, K_{ij} > 0.5, \end{cases} \quad (10)$$

where $i, j = 1, \dots, (l + u)$ and $S^c \in R^{(l+u) \times (l+u)}$. For l labeled data, reserve their class labels to form labeled discrimination matrix S^l as follows:

$$S_{ij}^l = \begin{cases} 1, & \text{if } y_i = y_j \\ -1, & \text{if } y_i \neq y_j, \end{cases} \quad (11)$$

where $i, j = 1, \dots, l$. Final discrimination matrix $S \in R^{(l+u) \times (l+u)}$ is built by combining clustering discrimination matrix and labeled discrimination matrix together:

$$S_{ij} = \begin{cases} S_{ij}^l, & \text{if } x_i, x_j \text{ are both labeled data} \\ S_{ij}^c, & \text{else,} \end{cases} \quad (12)$$

where $i, j = 1, \dots, (l+u)$. In summary, S_{ij} is 1 in two situations: firstly when i th instance and j th instance belong to the same class for labeled data or the same clustering for unlabeled data and secondly when the reliability of clustering is low.

Further, the optimal solution of classification should possess twinning constraints regularization containing lower intracluster compactness and higher intercluster separability as follows:

$$\min_f \frac{1}{2} \sum_{i,j=1}^{l+u} \|f(x_i) - f(x_j)\| W_{c,ij} - \frac{1}{2} \sum_{i,j=1}^{l+u} \|f(x_i) - f(x_j)\| W_{s,ij} \quad (13)$$

where W_c is weight matrix for intracluster in which $W_{c,ij}$ is 1 when S_{ij} is 1 and $W_{c,ij}$ is 0 when S_{ij} is -1 . W_s is weight matrix for intercluster in which $W_{s,ij}$ is 1 when S_{ij} is -1 and $W_{s,ij}$ is 0 when S_{ij} is 1.

Finally, the proposed framework utilizes cluster assumption; that is, data in the same cluster with high similarity weighted by clustering reliability should share the same class label or otherwise possess different class labels. By integrating clustering discrimination of labeled and unlabeled data with twinning constraints regularization described as (13), formulate optimization problem of the proposed CDMR framework as follows:

$$\min_f \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f(x_i)) + \gamma_A \|f\|_k^2 + \frac{\gamma_I}{2(l+u)^2} \sum_{i,j=1}^{l+u} W_{ij} (f(x_i) - S_{ij} f(x_j))^2 \quad (14)$$

$$+ \frac{1}{2} \sum_{i,j=1}^{l+u} \|f(x_i) - f(x_j)\| W_{c,ij} - \frac{1}{2} \sum_{i,j=1}^{l+u} \|f(x_i) - f(x_j)\| W_{s,ij}$$

$$W_{ij} = \begin{cases} W_{ij}^0, & \text{if } S_{ij} = 1 \\ K_{ij} W_{ij}^0, & \text{else,} \end{cases} \quad (15)$$

where W is the weight matrix of the whole data and W_{ij}^0 represents the similarity between instance x_i and instance x_j according to the distance between them in fuzzy clustering manifold structure and $S \in R^{(l+u) \times (l+u)}$ is the final discrimination matrix which integrates fuzzy clustering discrimination with labeled discrimination. In (14), the regularization

term $(\gamma_I/2(l+u)^2) \sum_{i,j=1}^{l+u} W_{ij}(f(x_i) - S_{ij}f(x_j))^2$ represents the fuzzy clustering discrimination of both labeled data and unlabeled data.

3.2. *CDMR Framework Based ELM (CDMR-ELM)*. Based on ELM and the proposed CDMR framework, we construct semisupervised classification model on the basis of CDMR-ELM. Substitute (1) into (14) to obtain the objective function as follows:

$$\begin{aligned} \operatorname{argmin}_{\beta} & \left\{ \frac{1}{l} \|H\beta - Y\|^2 + \gamma_A \beta H H^T \beta^T \right. \\ & + \frac{\gamma_I}{2(l+u)^2} \beta H H^T L_D H H^T \beta^T \\ & + \frac{\gamma_D}{2} \beta H H^T L_C H H^T \beta^T \\ & \left. - \frac{(1-\gamma_D)}{2} \beta H H^T L_S H H^T \beta^T \right\}, \end{aligned} \quad (16)$$

where $(1/l)\|H\beta - Y\|^2$ is square error of l labeled data, $L_D = D - W \circ S$ is a Laplacian matrix that is based on clustering discrimination of the whole dataset, $L_C = D_C - W_C$ is a Laplacian matrix for intracluster where D_C is a diagonal matrix denoted by $D_{C,ii} = \sum_{i,j=1}^{l+u} W_{C,ij}$, and $L_S = D_S - W_S$ is a Laplacian matrix for intercluster where D_S is a diagonal matrix denoted by $D_{S,ii} = \sum_{i,j=1}^{l+u} W_{S,ij}$.

By zeroing the gradient of the objective function with respect to β , convert (16) as follows:

$$\begin{aligned} \frac{1}{l} H (H\beta - Y)^T & + \left[\gamma_A H H^T + \frac{\gamma_I}{2(l+u)^2} H H^T L_D H H^T \right. \\ & \left. + \frac{\gamma_D}{2} H H^T L_C H H^T - \frac{(1-\gamma_D)}{2} H H^T L_S H H^T \right] \beta^T \\ & = 0. \end{aligned} \quad (17)$$

Then, the solution of the CDMR-ELM is obtained:

$$\begin{aligned} \beta^* & = H^T \left[H H^T + \gamma_A I + \frac{\gamma_I}{2(l+u)^2} L_D H H^T \right. \\ & \left. + \frac{\gamma_D}{2} L_C H H^T - \frac{(1-\gamma_D)}{2} L_S H H^T \right]^{-1} T, \end{aligned} \quad (18)$$

where I is the identity matrix with dimension of $l+u$. According to (1) and (2), the decision function of the proposed semisupervised classification model with regard to input x is shown as follows:

$$\begin{aligned} f(x) & = h(x) \beta^* = H^T \left[H H^T + \gamma_A I \right. \\ & + \frac{\gamma_I}{2(l+u)^2} L_D H H^T + \frac{\gamma_D}{2} L_C H H^T \\ & \left. - \frac{(1-\gamma_D)}{2} L_S H H^T \right]^{-1} T. \end{aligned} \quad (19)$$

Given the hidden weights a , biases b , and trade-off parameters γ_A , γ_I , and γ_D previously, the MSE of classification described as follows should be minimized to improve accuracy:

$$\operatorname{MSE}(a, b, \gamma_A, \gamma_I, \gamma_D) = \frac{1}{l+u} \sum_s^{l+u} (y_s - f(x_s))^2, \quad (20)$$

where y_s represents the predicted output and $f(x_s)$ represents the actual output for input data x_s .

4. The Optimized Classifier (MOFOA-CDMR-ELM)

4.1. *The Multiobjective Optimization Problem and Solutions*. Considering that the number of hidden layer nodes strongly influences the semisupervised classification efficiency and training time, the multi-objective optimization problem is to find the optimal SLFNs with a lower MSE and a smaller number of hidden nodes L simultaneously as follows:

$$\begin{aligned} \min & \quad (\operatorname{MSE}(a, b, \gamma_A, \gamma_I, \gamma_D), L), \\ \text{s.t.} & \quad (a, b) \in R^{L \times (d+1)}. \end{aligned} \quad (21)$$

The solutions of this multiobjective optimization problem are represented as follows:

$$\begin{aligned} (a, b, \gamma_A, \gamma_I, \gamma_D, L) & = (a_{11}, \dots, a_{d1}, a_{12}, \dots, a_{d2}, \dots, a_{1L}, \\ & \dots, a_{dL}, b_1, \dots, b_L, \gamma_A, \gamma_I, \gamma_D, L). \end{aligned} \quad (22)$$

Parameters γ_A , γ_I , and γ_D can control the reliability of the clustering discrimination from the semisupervised clustering method. If the values of these parameters are larger, the fuzzy clustering discrimination is more important. Otherwise, if the values of these parameters are small, CDMR will degenerate to smoothness assumptions as manifold regularization [12]. Therefore, the values of parameters γ_A , γ_I , and γ_D should be optimized with the aim of achieving better classification accuracy.

Unlike traditional single-objective optimization problem, optimization problem with multiobjective is impossible to find single solution which simultaneously minimizes all objectives [20–22]. This paper looks for a set of optimal solutions where there is no other efficient solution which improves one element of objectives without deteriorating the remaining elements.

4.2. *MOFOA-CDMR-ELM Classifier*. Considering FOA has possibility of sinking into local extremum and prematurity [15], this paper improves traditional FOA in the following two aspects:

- (1) Employing MSE to evaluate fitness function as follows:

$$\operatorname{Smell}_i = \operatorname{MSE}(X_i). \quad (23)$$

```

(1) Build the initial fruit fly swarm  $X^1$  in which each individual is in the form of (22).
(2) Evaluate the fitness function on training set by using (20).
(3) Set add and reduce variables.
    for  $k = 1$  to  $K$  do
(4) According to add and reduce, adjust  $L$  of each individual in  $X^k$ .
(5) Evaluate the new swarm  $X^{k'}$  on training set by using (20).
(6) for  $i = 1$  to max_it do
(7) for each individual in the  $X^{k'}$ 
(8) Adjust  $(a, b, \gamma_A, \gamma_I, \gamma_D)$  by using adaptively reduced search area by using (24)
(9) if MSE of new individual is better than previous one then
(10) New individual replaces previous one
(11) Reset add and reduce variables
(12) Reserve size_of_solution global optimal solutions  $X^*$  in population  $X^K$ .

```

ALGORITHM 1

```

(1) if add == 1 and reduce == 0 then
(2) new_L = L + Integer( $\omega * \min\{0.25 * (L_{mid} + L_{max} - L), L' - L\}$ )
(3) else if add == 0 and reduce == 1 then
(4) new_L = L - Integer( $\omega * \min\{0.25 * (L_{mid} + L - L_{min}), L - 1\}$ )
(5) else if add = reduce == 1 then
(6) new_L = arg_L min(MSE)
(7) else if add = reduce == 0 then
(8) new_L = L

```

ALGORITHM 2

(2) Adopting adaptively reduced search area for decision variable (a, b) along with iteration going on as follows:

$$\begin{aligned}
 \text{new_}a &= a \pm r_a(k), \\
 \text{new_}b &= b \pm r_b(k), \\
 r_{a,b}(k) &= \left(\frac{K-k}{K}\right)^2 * r_{\max},
 \end{aligned} \tag{24}$$

where K is the number of iterations, $r_{a,b}(k)$ represents the adaptive search area for the k th iteration, $k \in [1, K]$ is current iteration index, and r_{\max} is the maximum search area set as $1/2$ which is quarter of gap between high limit and low limit of a and b .

The algorithm starts with initializing fruit fly swarm X^1 consisting of *size_of_swarm* individuals represented as vector in (22) in which (a, b) are randomly assigned from uniform distribution between -1 and 1 , $(\gamma_A, \gamma_I, \gamma_D)$ are limited in the range of $(2^{-24}, 2^{24})$, and L is between 1 and the upper limit for hidden nodes. Next, introduce two variables *add* and *reduce* to control the search of optimal L by means of relationship between L and MSE. After adjustment of appropriate L , evaluate the new solutions and implement *max_it* times inner loop on them for adjusting parameters $(a, b, \gamma_A, \gamma_I, \text{ and } \gamma_D)$ in which three trade-off parameters are tuning during the range. Finally, reset the value of *add* and *reduce*. The main loop is repeated K times to search global optimal swarm X^* in X^K .

The MOFOA for optimizing CDMR-ELM is described as shown in Algorithm 1.

In this paper, we suppose relationship between the MSE and L is parabolic or linear. If L is proportional to MSE, set *add* to be 0 and set *reduce* to be 1 . If L is inversely proportional to MSE, set *add* to be 1 and set *reduce* to be 0 . If MSE does not improve by increasing or decreasing nodes, set both *add* and *reduce* to be 1 . If MSE decreases when L both increases and decreases, set both *add* and *reduce* to be 0 . Variables *add* and *reduce* guide the search of L as shown in Algorithm 2.

In Algorithm 2 L_{\max} , L_{\min} , and L_{mid} are maximum, minimum, and middle values of L in population X^k , ω is uniform random value in $(0, 1)$, and L' is the upper limit for hidden nodes.

5. Experiment Results and Discussion

5.1. Datasets and Experiment Setup. In order to evaluate the accuracy and efficiency of the proposed MOFOA-CDMR-ELM classifier, we perform a set of experiments on several real-world datasets from the UCI machine learning repository and benchmark repository frequently used for semisupervised learning [23]. The details of datasets are shown in Table 1.

Comparison experiments are implemented on two types of classifiers: one type is supervised classifier including SVM and ELM; the other type is semisupervised classifier including SSL-ELM [5, 24], LapRLS [7], LapSVM [8], and

TABLE 1: Details of the selected datasets for semisupervised learning.

Dataset	Classes	Attributes	Size
COIL2	2	1024	1440
COIL20	20	1024	1440
Shuttle	2	9	43500
USPST	10	256	2007
EEG Eye State	2	14	14980
EMGPA	11	8	10000
Seeds	3	7	210
Vertebral Column	2	6	310

TABLE 2: Parameters setting.

Parameter	Value	Description
<i>size_of_swarm</i>	20	Number of individuals in swarm
<i>K</i>	20	Number of outer iterations
<i>max_it</i>	100	Number of inner iterations
<i>L'</i>	100	Maximal number of hidden nodes

the proposed classifier. Divide each dataset into three subsets: testing set, validation set, and training set which is further partitioned into fixed labeled set and unlabeled set. Make sure the labeled set contains at least one sample of each class. Training set is used to train classifiers. Validation set containing labeled data is utilized for optimal model selection. Testing set is used to verify the classifier performance and efficiency. All experiments are implemented in MATLAB 7.0 which is running on a PC with CPU of 3.4 GHZ and RAM of 4.0 GB.

5.2. Parameters Setting. For ELM and SSL-ELM, adopt Gaussian function $\exp(-b\|x - a\|^2)$ and use grid search method to find out optimal trade-off parameter C between $\{2^{-20}, 2^{-19}, \dots, 2^{19}, 2^{20}\}$ and number of hidden nodes L between $\{10, 20, \dots, N\}$ where N is the size of each dataset. For classifiers based on SVM, search optimal parameter C between $\{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$ according to classification accuracy. Furthermore, classifiers based on SVM adopt one-to-rest method to solve multiclass classification problem. The optimal weights γ_A and γ_I of regularization items of LapSVM and LapRLS are searched from grid $\{10^{-5}, 10^{-4}, \dots, 10^5, 10^6\}$ by cross validation. The setting of parameters required in the proposed classifier is shown in Table 2.

5.3. Effectiveness of the Proposed MOFOA. In order to evaluate the effectiveness of the proposed optimization method in searching for optimal parameters as (23), we compare three classifiers containing the proposed CDMR-ELM, FOA-CDMR-ELM, and MOFOA-CDMR-ELM. Classifier CDMR-ELM with random (a_i, b_i) can obtain optimal trade-off parameters γ_A , γ_I , and γ_D by implementing 10-fold cross validation on validation set for 100 times. FOA-CDMR-ELM employs classification error rate to guide the search of optimal weights and biases (a_i, b_i) as well as weight of regularization items γ_A , γ_I , and γ_D by giving a fixed number of hidden nodes

L . MOFOA-CDMR-ELM searches appropriate number of hidden nodes, weights, and biases (a_i, b_i) as well as weight of regularization items γ_A , γ_I , and γ_D by simultaneously minimizing MSE and the number of hidden nodes L . Table 3 shows the mean value of classification accuracy and number of hidden nodes by three classifiers on all datasets. Data in bold type represent the optimal classification result and hidden nodes.

From Table 3, we can see that the proposed MOFOA-CDMR-ELM classifier is better than the other two competitive classifiers in 75%. This result fully verifies the effectiveness of the proposed optimization method based on FOA since it adopts adaptively reduced search area for searching in iteration to reduce possibility of sinking into local extremum and premature. Further, focusing on minimizing both MSE and hidden nodes, MOFOA can obtain superior networks with less hidden nodes under the guarantee of better accuracy.

5.4. Comparison of Performance. We compare classification accuracy between some state-of-the-art supervised classifiers and semisupervised classifiers on above-mentioned datasets to evaluate efficiency and effectiveness of the proposed classifier. Table 4 shows the mean value and standard deviation of classification accuracy and Table 5 shows the mean value and standard deviation of running time of all the compared classifiers on 8 datasets.

From Table 4, we can conclude the following:

- (1) LapRLS and LapSVM outperform supervised classifiers SVM and ELM in semisupervised learning even with a few labeled data, since LapRLS and LapSVM adopt manifold regularization to utilize unlabeled data according to nonlinear geometrical manifold structure embedding in the whole data.
- (2) Among three existing semisupervised classifiers, by constructing a framework that integrates manifold assumption with constraints between all the labeled data to relieve misclassification in boundary area and enhance the smoothness of decision function, SSL-ELM obtains better classification accuracy than LapRLS and LapSVM.
- (3) The proposed classifier outperforms SSL-ELM especially on multiclass datasets since it adopts unsupervised fuzzy clustering method and considers inner cluster and intercluster constraints not only between labeled data but also between unlabeled data. Further, the proposed MOFOA plays an important role in enhancing the performance by searching for optimal parameters.

From Table 5, we can see that training time of SVM and ELM is obviously less than semisupervised classifier especially on dataset with large size since they are trained only based on labeled data. To be fair, comparing four semisupervised classifiers trained based on both labeled and unlabeled data, training time of LapSVM classifier for multiclass dataset is more than others. It is possibly due to the fact that one-to-rest method seriously increases running

TABLE 3: Comparisons of average classification accuracy on dataset.

Dataset	CDMR-ELM		FOA-CDMR-ELM		MOFOA-CDMR-ELM	
	Accuracy (%)	Hidden nodes	Accuracy (%)	Hidden nodes	Accuracy (%)	Hidden nodes
COIL2	91.42	34.5	91.86	31.7	92.41	22.1
COIL20	91.59	39.2	92.23	30.5	94.57	25.0
Shuttle	92.88	100	93.46	84.7	95.65	86.9
USPST	91.45	46.9	92.25	79.9	92.59	31.4
EEG Eye State	89.58	100	88.62	87.6	89.15	89.2
EMGPA	89.78	97.7	91.48	89.5	90.18	76.5
Seeds	92.33	29.9	93.21	22.4	96.88	18.3
Vertebral Column	90.45	27.4	91.65	19.7	93.76	12.8

TABLE 4: Mean value and standard deviation of classification accuracy.

Dataset	SVM	ELM	LapRLS	LapSVM	SSL-ELM	The proposed classifier
	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)	Accuracy (%)
COIL2	82.11 (± 2.33)	83.25 (± 2.02)	88.87 (± 1.70)	88.52 (± 1.45)	90.15 (± 1.30)	92.68 (± 2.22)
COIL20	81.38 (± 1.06)	82.88 (± 2.34)	87.25 (± 2.10)	87.75 (± 1.02)	91.38 (± 1.50)	93.15 (± 1.67)
Shuttle	85.28 (± 1.85)	85.94 (± 1.68)	91.05 (± 2.55)	91.25 (± 2.04)	92.18 (± 2.97)	95.60 (± 2.01)
USPST	81.45 (± 2.95)	81.98 (± 1.91)	90.12 (± 2.33)	90.38 (± 2.19)	91.06 (± 2.22)	93.48 (± 1.59)
EEG Eye State	73.89 (± 2.35)	75.13 (± 2.58)	85.67 (± 1.85)	85.22 (± 1.88)	87.50 (± 2.44)	89.55 (± 2.10)
EMGPA	80.44 (± 1.87)	81.26 (± 2.51)	86.85 (± 2.02)	87.30 (± 1.52)	88.10 (± 1.68)	91.68 (± 2.05)
Seeds	78.81 (± 2.09)	79.63 (± 2.65)	86.92 (± 2.01)	87.56 (± 2.33)	90.25 (± 1.32)	96.65 (± 1.72)
Vertebral Column	82.12 (± 1.55)	83.03 (± 1.99)	84.68 (± 1.96)	84.21 (± 1.19)	89.17 (± 1.95)	93.59 (± 1.87)

TABLE 5: Mean value and standard deviation of training time.

Dataset	SVM	ELM	LapRLS	LapSVM	SSL-ELM	The proposed classifier
	Training time (s)	Training time (s)	Training time (s)	Training time (s)	Training time (s)	Training time (s)
COIL2	$3.75 (\pm 0.12) \times 10^{-3}$	$1.08 (\pm 0.09) \times 10^{-3}$	1.68 (± 0.36)	1.55 (± 0.08)	0.37 (± 0.02)	0.13 (± 0.02)
COIL20	$3.35 (\pm 0.27) \times 10^{-2}$	$1.48 (\pm 0.27) \times 10^{-3}$	2.02 (± 0.19)	2.82 (± 0.17)	0.51 (± 0.08)	0.27 (± 0.03)
Shuttle	$4.56 (\pm 0.22) \times 10^{-3}$	$2.19 (\pm 0.10) \times 10^{-3}$	29.44 (± 1.51)	32.38 (± 2.90)	12.52 (± 3.67)	9.82 (± 1.66)
USPST	$3.09 (\pm 0.36) \times 10^{-3}$	$2.12 (\pm 0.22) \times 10^{-3}$	29.81 (± 2.75)	38.35 (± 0.20)	5.76 (± 0.25)	3.01 (± 0.28)
EEG Eye State	$3.78 (\pm 0.29) \times 10^{-2}$	$2.81 (\pm 0.25) \times 10^{-2}$	26.71 (± 3.20)	33.79 (± 1.95)	6.36 (± 0.96)	5.89 (± 0.77)
EMGPA	$4.81 (\pm 0.65) \times 10^{-2}$	$1.73 (\pm 0.05) \times 10^{-2}$	18.60 (± 0.98)	25.23 (± 1.65)	6.88 (± 0.63)	5.27 (± 0.49)
Seeds	$2.32 (\pm 0.16) \times 10^{-4}$	$5.82 (\pm 0.27) \times 10^{-5}$	$7.33 (\pm 0.35) \times 10^{-2}$	$8.09 (\pm 0.26) \times 10^{-2}$	$3.13 (\pm 0.22) \times 10^{-2}$	$2.95 (\pm 0.10) \times 10^{-2}$
Vertebral Column	$2.57 (\pm 0.11) \times 10^{-4}$	$6.32 (\pm 0.19) \times 10^{-5}$	$7.82 (\pm 0.25) \times 10^{-2}$	$9.32 (\pm 0.11) \times 10^{-2}$	$3.32 (\pm 0.13) \times 10^{-2}$	$3.75 (\pm 0.10) \times 10^{-2}$

time in iterative process. The proposed classifier optimized by MOFOA obtains optimal parameters in model with high classification and fewer hidden nodes which lead to fast learning speed according to the theory of ELM that the number of hidden nodes is proportional to training time. In general, the proposed classifier can achieve better performance with optimal learning speed.

5.5. Performance with Different Number of Labeled and Unlabeled Data. The previous experiments are implemented under fixed labeled set and unlabeled set. If the number of labeled and unlabeled data varies gradually, the performance of classifiers exhibits some change tendency. Figure 1 shows the performance variation of ELM, LapRLS, LapSVM, SSL-ELM, and the proposed classifier on two representative

datasets, Shuttle and Seeds, with different number of labeled data by varying proportion of labeled data and unlabeled data in training set. Figure 2 shows the performance variation of these classifiers with different number of unlabeled data.

From Figure 1, we can observe that, with the increase of number of labeled data, the classification accuracy of every classifier is stably improved. Further, accuracy of the proposed classifier outperforms others all along. From Figure 2, we can see that, with the increase of number of unlabeled data, the classification accuracy of ELM is maintained unchanged since it works only based on labeled data while accuracy of the other semisupervised classifier is enhanced obviously. Further, even with very few unlabeled data, the proposed classifier outperforms SSL-ELM, LapRLS, and LapSVM because it constructs manifold structure by

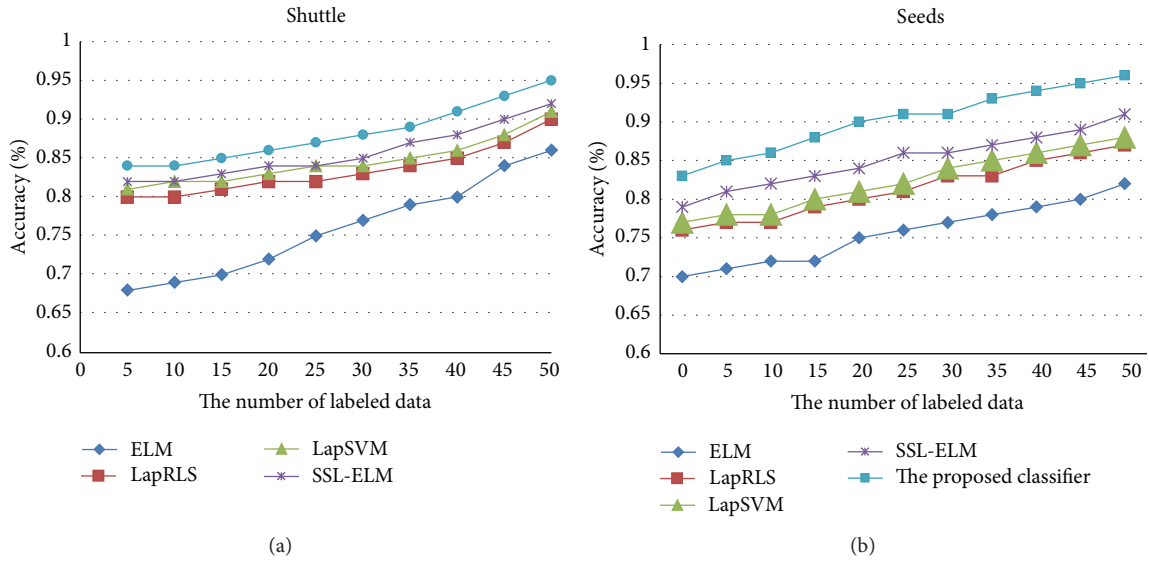


FIGURE 1: Classification accuracy with respect to different labeled data.

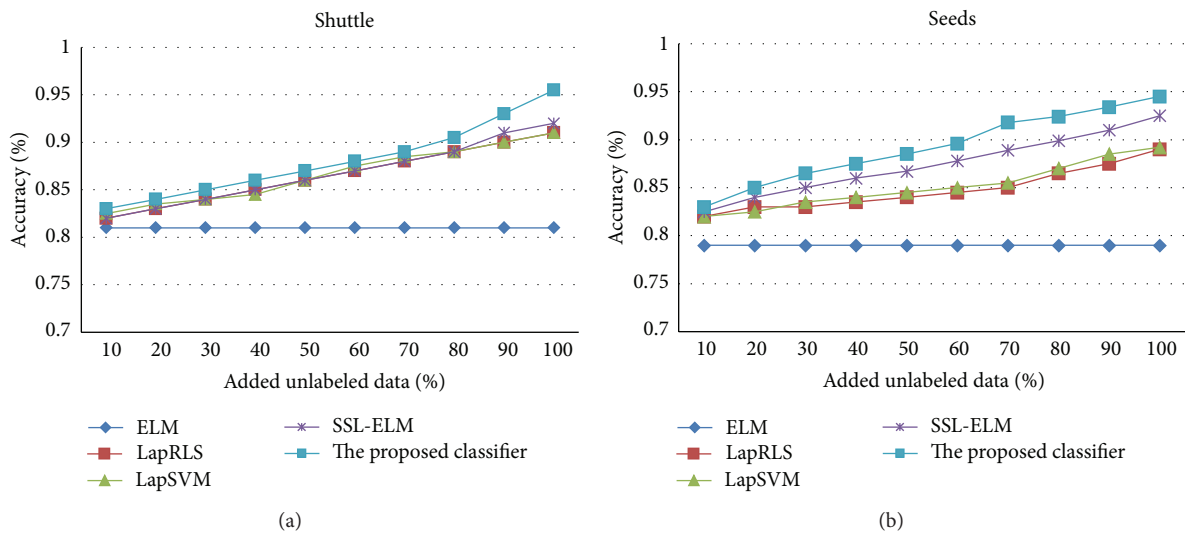


FIGURE 2: Classification accuracy with respect to different unlabeled data.

fully utilizing both unlabeled data and labeled data which is effective for supervised learning. In general, the results verify that the proposed classifier can obtain better performance in dynamic semisupervised classification since it integrates discrimination of both labeled and unlabeled data with twinning constraints of fuzzy clusters.

6. Conclusion

In this paper, we propose a feasible semisupervised learning method in terms of clustering discrimination of the whole data and twinning constraints regularization named CDMR. Further, we integrate ELM with the proposed semisupervised learning framework to achieve semisupervised classification. With the purpose of enhancing the classification accuracy

and training speed of the proposed classifier, we build a novel multiobjective FOA which simultaneously minimizes the number of hidden nodes and MSE to obtain optimal parameters of classifier to guarantee that there are no other SLFNs with higher accuracy and fewer or equal number of hidden nodes. Experiments' results on several datasets confirm the effectiveness and efficiency of the proposed MOFOA-CDMR-ELM classifier. In the future, we will deeply study the sparsity problem of matrix multiplication to further reduce training time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant no. 70701013, the Scientific Research and Technology Development Plan Project of Guangxi Province under Grant no. 2013F020202, and the Research Project of Liuzhou GM-Wuling Limited Liability Company under Grant no. 20132h0261. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

- [1] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.
- [2] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [3] R. A. Berk, "Support vector machines," in *Statistical Learning from a Regression Perspective*, Springer Series in Statistics, pp. 1–28, Springer, 2008.
- [4] G.-M. Lim, D.-M. Bae, and J.-H. Kim, "Fault diagnosis of rotating machine by thermography method on support vector machine," *Journal of Mechanical Science and Technology*, vol. 28, no. 8, pp. 2947–2952, 2014.
- [5] G. Huang, Sh. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2405–2417, 2014.
- [6] M. Belkin, V. Sindhwani, and P. Niyogi, "Manifold regularization: a geometric framework for learning from labeled and unlabeled examples," *Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.
- [7] X. Zhu and A. B. Goldberg, *Introduction to Semi-Supervised Learning*, Morgan & Claypool, San Rafael, Calif, USA, 2009.
- [8] S. Melacci and M. Belkin, "Laplacian support vector machines trained in the primal," *Journal of Machine Learning Research*, vol. 12, pp. 1149–1184, 2011.
- [9] W.-J. Chen, Y.-H. Shao, and N. Hong, "Laplacian smooth twin support vector machine for semi-supervised classification," *International Journal of Machine Learning and Cybernetics*, vol. 5, no. 3, pp. 459–468, 2014.
- [10] Y. Zhou, B. Liu, and S. Xia, "Semi-supervised extreme learning machine with manifold and pairwise constraints regularization," *Neurocomputing*, vol. 149, pp. 180–186, 2015.
- [11] F. Wu, W. Wang, Y. Yang, Y. Zhuang, and F. Nie, "Classification by semi-supervised discriminative regularization," *Neurocomputing*, vol. 73, no. 10–12, pp. 1641–1651, 2010.
- [12] Y. Wang, S. Chen, H. Xue, and Z. Fu, "Semi-supervised classification learning by discrimination-aware manifold regularization," *Neurocomputing*, vol. 147, pp. 299–306, 2014.
- [13] W.-T. Pan, "Using modified fruit fly optimisation algorithm to perform the function test and case studies," *Connection Science*, vol. 25, no. 2-3, pp. 151–160, 2013.
- [14] S.-M. Lin, "Analysis of service satisfaction in web auction logistics service using a combination of Fruit fly optimization algorithm and general regression neural network," *Neural Computing and Applications*, vol. 22, no. 3-4, pp. 783–791, 2013.
- [15] W. T. Pan, "A new evolutionary computation approach: fruit fly optimization algorithm," in *Proceedings of the Conference on Digital Technology and Innovation Management*, Taipei, Taiwan, 2011.
- [16] S. M. Mousavi, N. Alikar, and S. T. Akhavan Niaki, "An improved fruit fly optimization algorithm to solve the homogeneous fuzzy series-parallel redundancy allocation problem under discount strategies," *Soft Computing*, 2015.
- [17] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [18] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '04)*, pp. 985–990, Budapest, Hungary, July 2004.
- [19] S. Ghosh and S. K. Dubey, "Comparative analysis of K-means and fuzzy C-means algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 4, pp. 35–39, 2013.
- [20] D. Lahoz, B. Lacruz, and P. M. Mateo, "A bi-objective micro genetic extreme learning machine," in *Proceedings of the IEEE Workshop on Hybrid Intelligent Models and Applications (HIMA '11)*, pp. 68–75, IEEE, April 2011.
- [21] D. Lahoz, B. Lacruz, and P. M. Mateo, "A multi-objective micro genetic ELM algorithm," *Neurocomputing*, vol. 111, pp. 90–103, 2013.
- [22] C. Coello, List of references on evolutionary multi-objective optimization, 2011, <http://delta.cs.cinvestav.mx/~ccoello/EMOO/EMOObib.html>.
- [23] A. Asuncion and D. Newman, "UCI Machine Learning Repository," 2010, <http://archive.ics.uci.edu/ml>.
- [24] J. Liu, Y. Chen, M. Liu, and Z. Zhao, "SELM: semi-supervised ELM with application in sparse calibrated location estimation," *Neurocomputing*, vol. 74, no. 16, pp. 2566–2572, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

