Multimed Tools Appl
DOI 10.1007/s11042-017-4494-3

Hyperspectral remote sensing image classification with information discriminative extreme learning machine

Deqin Yan[1] · **Yonghe Chu**[1] · **Lina Li**[1] · **Deshan Liu**[1]

Received: 29 May 2016 / Revised: 1 January 2017 / Accepted: 8 February 2017

Abstract Hyperspectral remote sensing image classification is important aspect of current research. Extreme learning machine (ELM) has been widely used in the field of pattern recognition for its efficient and good generalization performance. With the study of hyperspectral remote sensing image classification, this paper proposes an information discriminative extreme learning machine (IELM). IELM inherits the advantages of ELM, can solve the problems that ELM learning is insufficient for hyperspectral remote sensing image with limited scale of sample data. The proposed algorithm is tested by experiments for hyperspectral remote sensing image classification. The experiment results show that the proposed algorithm has better classification effect.

Keywords Extreme learning machine · Pattern recognition · Hyperspectral remote sensing image · Discrimination information

1 Introduction

Over the last decade, with the development of image processing technology, hyperspectral images have been widely studied. In the field of hyperspectral image research, the most important task is to classify the images. However, there are many challenges in hyperspectral remote sensing image classification problem, such as the imbalance and high dimension among the limited training samples. Hyperspectral remote sensing images have complex geometric structure, have high computational complexity at the time of classification, and so on. In recent years, in order to overcome these problems and achieve good results, machine learning methods have been widely used in hyperspectral image classification, such as artificial neural networks (ANNs) [30],

✉ Deqin Yan
yandeqin@163.com

[1] School of Computer and Information Technology, Liaoning Normal University, Dalian 116081, China

Published online: 20 March 2017

support vector machine (SVM) [4–6, 8, 20], multinomial logistic regression (MLR) [14], active learning (AL) [7, 25], semi-supervised learning (SSL) [12], Manifold learning (ML) [18, 36] and so on. More methods have also been widely used in hyperspectral remote sensing image classification, such as sparse representation [3], spectral clustering [34]. On the issue of remote sensing image classification, Belkacem Baassou et al. used Spatial Pixel Association (SPA) processing method to further improve the classification performance of SVM techniques [1]. Tao et al. proposed dual-regularized KISS (DR-KISS) metric learning [33]. Lu et al. used the joint dictionary coefficient coding method to realize multispectral detection [17]. In [32], Tao et al. proposed a novel Hessian regularized support vector machine algorithm on image annotation. In [38], Yuan et al. used multitask joint sparse representation (MJSR) and a stepwise Markov random filed framework to improve the accuracy and robustness of remote sensing image classification. Aiming to solve the problem of Salient Band Selection with hyperspectral remote sensing image, Wang et al. proposed an solution based on manifold ranking [36]. However, due to the high dimensionality and complexity of hyperspectral image in spectral domain, finding optimal parameters for parametric supervised algorithms is always time-consuming and difficult. It is still a critical problem to achieve high-performance classification with fast speed and high efficiency.

Recently, Huang et al. [11, 31, 35] proposed extreme learning machine (ELM) [26] based on the structure of single hidden layer feedforward neural networks (SLFNs). ELM converted the solving process of traditional neural networks into a linear model in which only output weights is determined, input weights and offset value of hidden nodes are randomly generated. The algorithm of ELM can avoid possibility of that the traditional neural network learning method is with slow convergence speed and local optimal solution. Because of the good generalization ability of ELM, it can be applied to many areas [16, 19, 23, 27, 37, 39, 41]. In the field of hyperspectral remote sensing image research, Pal et al. applied ELM to land cover classification [21, 22], compared with BP [28] neural networks and SVM, ELM achieved better classification results, and its computational complexity is much smaller than the BP and SVM. Bazi et al. used a differential evolution method to select the optimal parameters in the kernel-based ELM for HSI classification [2]. Samat et al. proposed two ensemble ELM methods based on Bagging and AdaBoost to improve the stability of ELM [29]. Heras et al. explored two ELM-based spatial–spectral classifiers for HSI classification using watershed transform and spatially regularization methods [9].

Traditional ELM and related algorithms only can do the supervised learning classification work for hyperspectral image on limited scale samples. In the process of classification the ELM and related algorithms only learn the feature of single sample, so the geometrical characteristic and discriminate information between samples are neglected, which limited the generalization performance of ELM. As we know, there are classification attributes and distribution characteristics with samples. Exploiting this information is important to enhance generalization performance of ELM. Thus the geometric characteristics and the discrimination information of the data samples play an important role on the classification performance ELM. Based on the analysis, combined with the characteristics of hyperspectral remote sensing image, we proposed a kind of discriminative information regularized extreme learning machine (IELM). IELM takes into account the geometrical characteristics and the discrimination information of data, by maximizing inter-class scatter degree and minimizing within-class scatter degree, optimizing the output weights

to advance learning and classification performance of extreme learning machine. The advantages of IELM are:

1) It inherits the advantages of ELM, to some extent, avoids the problem of insufficient learning of ELM.
2) It introduces the intra-class and within-class scatter degree into the ELM, and it not only takes advantage of the difference of discrimination information between data samples, but also avoids the phenomenon that information of data sample is overlapping.
3) By using MMC [13] method, it effectively solves problem of maximize intra-class and minimize within-class scatter matrix.

In order to evaluate and verify the performance of proposed method, we compared IELM with ELM, SVM, CRNN (Collaborative representation nearest neighbor classifier, CRNN) [15] and MSMRF (Hyperspectral Image Classification via Multitask Joint Sparse Representation and Stepwise MRF Optimization) [38] by experiments with two kinds hyperspectral remote sensing image data: Indian Pines, Salinas scene two hyperspectral remote sensing image data. Experimental results show that the proposed method can achieve better classification results. The formula definitions are listed in (Table 1).

Table 1 List of formulas

Formula abbreviation	Formula definition
X	$X=(x_1, x_2, \cdots, x_N)^T \in R^{D\times N}$
a_i	$a_i=(a_{i1}, a_{i2}, \cdots, a_{in})$
β_i	$\beta_i=(\beta_{i1}, \beta_{i2}, \cdots, \beta_{im})$
$h(x_i)$	$h(x_i)=(g(a_1x_i+b_1), g(a_2x_i+b_2), \cdots, g(a_Lx_i+b_L))$
S	$S=S_B-(1-n)S_W(0\le n\le 1)$
α_i	$\alpha_i=[\alpha_{i1}, \cdots \alpha_{im}]^T$
α	$\alpha=[\alpha_1, \cdots \alpha_N]$
t	$t_j=(t_{j1}, t_{j2}, \cdots, t_{jm})^T \in R^m$
ξ	$\xi_i=(\xi_{i1}, \cdots, \xi_{1m})^T$
S_B	$S_b=\sum_{i=1}^{C}\sum_{j=1}^{N_C}\left(x_j-u^i\right)L_i\left(x_j-u^i\right)^T$
S_W	$S_W=\sum_{i=1}^{C}(u^i-u)(u^i-u)^T$
H	$H=\begin{pmatrix} g(a_1\cdot x_1+b_1) & \cdots & g(a_L\cdot x_1+b_L) \\ \vdots & \ddots & \vdots \\ g(a_1\cdot x_N+b_1) & \cdots & g(a_L\cdot x_N+b_L)\end{pmatrix}_{N\times L}$
F	$F=\begin{pmatrix} f(a_1\cdot x_1+b_1) & \cdots & f(a_L\cdot x_1+b_L) \\ \vdots & \ddots & \vdots \\ f(a_1\cdot x_N+b_1) & \cdots & f(a_L\cdot x_N+b_L)\end{pmatrix}_{N\times L}$
T	$T=\begin{pmatrix} t_1^T \\ \vdots \\ t_N^T\end{pmatrix}=\begin{pmatrix} t_{11\cdots}, t_{1m} \\ \vdots \\ t_{N1}\cdots t_{Nm}\end{pmatrix}$
β	$\beta=\begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T\end{pmatrix}_{L\times m}$

2 ELM

Given N different samples (x_j, t_j), it can be expressed as $X=(x_1, x_2, \cdots, x_N)^T \in R^{D\times N}$, where $t_j=(t_{j1}, t_{j2}, \cdots, t_{jm})^T \in R^m$, the ELM learning form can be expressed as:

$$\sum_{i=1}^{L} \beta_i g\left(a_i \cdot x_j + b_i\right) = t_j \tag{1}$$

Where $j=1, 2, \cdots, N, a_i=(a_{i1}, a_{i2}, \cdots, a_{in})$ is input weight value vector which connect the $i-th$ hidden node with the input nodes, $\beta_i=(\beta_{i1}, \beta_{i2}, \cdots, \beta_{im})$ is output weight value vector which connect the $i-th$ hidden node the output nodes, x_j is offset value of the $i-th$ hidden node, represents the inner product of a_i and x_j, $t_j=(t_{j1}, t_{j2}, \cdots, t_{jm})^T \in R^m$ is the desired output vector of the sample H. Eq. (1) can be rewritten as follows:

$$H\beta = T \tag{2}$$

Where H is the output matrix of hidden layer nodes, β is a network hidden layer output matrix, T is the desired output matrix:

$$H = \begin{pmatrix} g(a_1 \cdot x_1 + b_1) & \cdots & g(a_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(a_1 \cdot x_N + b_1) & \cdots & g(a_L \cdot x_N + b_L) \end{pmatrix}_{N\times L} \tag{3}$$

$$\beta = \begin{pmatrix} \beta_1{}^T \\ \vdots \\ \beta_L{}^T \end{pmatrix}_{L\times m,} \quad T = \begin{pmatrix} t_1{}^T \\ \vdots \\ t_N{}^T \end{pmatrix} = \begin{pmatrix} t_{11\cdots}, t_{1m} \\ \vdots \\ t_{N1}\cdots t_{Nm} \end{pmatrix} \tag{4}$$

When the hidden layer node number is equal to the number of training samples ($L=N$), we can directly use inverse matrix H to get the optimal output weighting matrix β through the Eq. (2). But in most cases of hidden nodes number is much less than the number of training samples ($L<<N$), so matrix H is singular, we use least squares method to solve Eq. (2):

$$\hat{\beta} = \arg\min_{\beta} \|H\beta - T\|^2 = H^+ T \tag{5}$$

Where H^+ is the generalized inverse matrix of H.

In order to improve the stability and generalization ability of traditional ELM, Huang proposed equality optimization constraints based ELM [10], it's formula not only to minimize the training error ξ, but also minimize output weights β, and therefore target formula of equality optimization constraints based ELM can be written as:

$$\begin{aligned} &\min_{W} \frac{1}{2}\|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^{N} \xi_i{}^2 \\ &s.t. h(x_i)\beta = t_i{}^T - \xi_i{}^T \\ &i = 1, 2, \cdots, N \end{aligned} \tag{6}$$

In Eq. (6) $h(x_i) = (g(a_1 x_i + b_1), g(a_2 x_i + b_2), \cdots, g(a_L x_i + b_L))$, $\xi_i = (\xi_{i1}, \cdots, \xi_{1m})^T$ corresponding to the training error vector of the sample x_i, C is the penalty parameter. The optimization model of formula (6) can be solved by Lagrange multiplier method. It is transform into unconditional optimization problem to calculate:

$$\beta = H^T \left({}^{I}/_{C} + HH^T \right)^{-1} T \tag{7}$$

By the Eq. (7) we can be obtain output function of ELM:

$$g(x) = h(x)\beta = h(x)H^T \left(\frac{I}{C} + HH^T \right)^{-1} T \tag{8}$$

ELM algorithm can be summarized as:

1) Initialize the training sample set;
2) Randomly assign network input weights a_i and offset values b_i, $i = 1, 2, \cdots, L$;
3) Calculate the hidden layer nodes output matrix H by activation function;
4) Calculate the output weights, $\beta = H^T \left({}^{I}/_{C} + HH^T \right)^{-1} T$.

3 IELM

Traditional ELM and related algorithms neglect the correlation of data which can reduce the learning ability of ELM. Aiming to exploit the geometrical characteristic and discriminate information of data, we proposed a kind of discriminative information regularized extreme learning machine (IELM) as following:

Given N different samples (x_j, t_j), intra-class scatter matrix S_B, within-class scatter matrix S_W can be expressed as:

$$S_b = \sum_{i=1}^{C} \sum_{j=1}^{N_C} \left(x_j - u^i \right) L_i \left(x_j - u^i \right)^T \tag{9}$$

$$S_W = \sum_{i=1}^{C} \left(u^i - u \right) \left(u^i - u \right)^T \tag{10}$$

Where C is the number of category of data sample, L_i is the Laplacian matrix of i -th class, u^i of i -th class, u is overall mean of data sample. Inter-class scatter matrix and within-class scatter matrix respectively embody distribution characteristics and discrimination information of the input data sample space.

Definition 1 (information difference matrix) matrix $S = S_B - (1 - n)S_W (0 \le n \le 1)$ is defined as formation difference matrix. Where n is a non-negative constant.

In definition 1, parameter n is a parameter to adjust discrimination information of between the class and inter-class. The optimization scheme of IELM can be described as:

$$\begin{aligned} &\min_{W} tr\left(\beta^T S \beta \right) + \frac{1}{2} C \sum_{i=1}^{N} \varepsilon^2 \\ &s.t. \sum_{j=1}^{L} \beta_i f\left(a_j \cdot x_i + b_j \right) = t_i^T - \varepsilon_i^T \\ &i = 1, 2, \cdots, N \end{aligned} \tag{11}$$

Where β is the output weight matrix, ε is training error, C is the penalty parameter, $f(x)$ is activation function. The corresponding Lagrange function of formula (11) is:

$$L = tr\left(\beta^T S\beta\right) + \frac{1}{2}C\sum_{i=1}^{N}\varepsilon^2 + \sum_{i=1}^{N}\sum_{j=1}^{m}\alpha_{ij}\left(f(x_i)\beta_j - t_{ij} + \varepsilon_{ij}\right) \tag{12}$$

Where β_j is the weight of output node of $j-th$ $\beta = [\beta_1, \beta_2, \cdots, \beta_m]$ according to KKT conditions:

$$\frac{\partial L}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum_{i=1}^{N}\alpha_{ij} f(x_i)^T \rightarrow \beta S = F^T\alpha \tag{13}$$

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \rightarrow \alpha_i = C\varepsilon_i, i = 1, \cdots, N \tag{14}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow f\left(x_i\right)^i \beta - t_i{}^T + \varepsilon_i{}^T = 0 \tag{15}$$

Where $\alpha_i = [\alpha_{i1}, \cdots \alpha_{im}]^T, \alpha = [\alpha_1, \cdots \alpha_N]$。

1) When the number of training samples is greater than the number of hidden layer nodes, namely $N >> L$, substitute formula (13) into formula (15), we can get:

$$\beta = \left(F^T F + \frac{S}{C}\right)^{-1} F^T T \tag{16}$$

From formula (16), the output function of IELM is:

$$g(x) = f(x)\beta = f(x)\left(F^T F + \frac{S}{C}\right)^{-1} F^T T \tag{17}$$

2) When the number of training samples is smaller than the number of hidden layer nodes, namely N<<L, substitute formula (13) and formula (14) into formula (15), we can get:

$$\left(F^T F + \frac{S}{C}\right)\alpha = ST \tag{18}$$

substitute formula (13) into formula (18), we can get:

$$\beta = F^T\left(F^T F + \frac{S}{C}\right)^{-1} ST \tag{19}$$

From formula (19), the output function of IELM is:

$$g(x) = f(x)\beta = f(x)F^{\mathrm{T}}\left(F^T F + \frac{S}{C}\right)^{-1} ST \tag{20}$$

IELM algorithm can be summarized as follows:

1) Initialize the training sample set, calculate S_B and S_W, calculated of S according to the definition 1;
2) Randomly assign network input weights a_i and offset values b_i, $i = 1, 2, \cdots, L$;
3) Calculate the output matrix F of hidden layer nodes by activation function;
4) Calculate the output weights, $\beta = F^T\left(F^T F + \frac{S}{C}\right)^{-1} ST$ or $\beta = \left(F^T F + \frac{S}{C}\right)^{-1} F^T T$.

4 Hyperspectral remote sensing image classification experiments

4.1 Hyperspectral datasets

In this section, we investigate the performance of the proposed IELM on three benchmark hyperspectral datasets, i.e., Indian Pines, and Salinas. The two datasets are public available hyperspectral datasets. The Indian Pines dataset is a classic benchmark to evaluate the performance of HSI classification algorithms because of the widespread presence of mixed pixels is in all available classes and there is unbalanced number of labeled pixels per class. The Salinas datasets have very high spatial resolution and large number of labeled pixels.

AVIRIS hyperspectral image of Indian Pines consists of 145 × 145 pixels by 220 bands of radiance data with about two-thirds agriculture, and one-third forest or other natural perennial vegetation. We have reduced the number of bands to 200 by removing bands covering the region of water absorption. The ground truth map in Fig. 1. Indiana Pines data set property settings are shown in Table 2.

Salinas scene data is data AVIRIS sensor collected in the Salinas Valley collection, which contains 204 bands, the size of the image is 512 × 217, with 3.7-m geometric resolution, the ground truth map in Fig. 2, Salinas scene feature dataset contains 16 categories, Salinas scene dataset property settings are shown in Table 3.

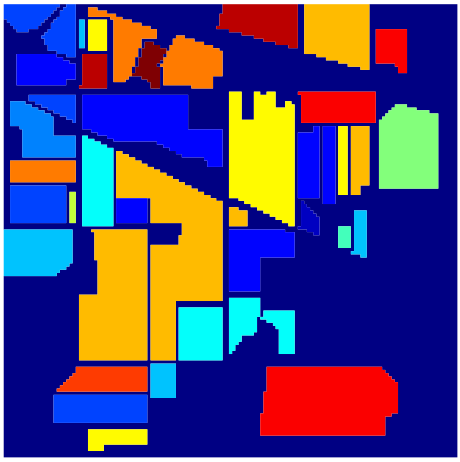

Fig. 1 Ground truth of Indian Pines

Table 2 Indian Pines category information data sets

Class	Samples
Alfalfa	54
Corn-notill	1434
Corn-mintill	834
Corn	234
Grass-pasture	497
Grass-trees	747
Grass-pasture-mowed	26
Hay-windrowed	489
Oats	20
Soybean-notill	968
Soybean-mintill	2468
Soybean-clean	614
Wheat	212
Woods	1294
Buildings-Grass-Trees-Drives	380
Stone-Steel-Towers	95
Total	10,336

4.2 Experiments and parameter setting

To evaluate the proposed algorithm, we compare IELM with the classical classifiers, such as SVM, general ELM, CRNN [15], and MSMRF [38] by experiments using the data sets described in section 4.1. For Indian Pines, Salinas scene data set the feature extraction algorithm of literature [40] is used to get the feature of each sample. Then the features are used as input data of classification algorithms (IELM, SVM, ELM, CRNN and MSMRF). 1% of labeled samples from each class are randomly chosen for training, and the remaining samples are used for testing. In this paper, different methods are compared based on the overall accuracy (OA), average accuracy (AA) and Kappa coefficient, where OA is the percentage of correctly classified pixels in the testing set, AA is the mean of class-specific accuracy values, and Kappa is the percentage of agreement corrected by the number of agreements that would be expected purely by chance [24]. All experiments are conducted using MATLAB R2015b on a computer with 2.80 GHz CPU and 8.0 GB RAM.

In the experiments, 10-fold cross-validation technique was used to obtain the optimal parameters. In all experiments, RBF kernel was used for SVM, in which the range of gamma factor is [0.01, 14], the range of cost factor is [0.01, 1024], and all involved parameters were optimized by using grid search method. The parameter set of CRNN and MSMRF is as in [15, 38]. Cosine function was used as the activation function in ELM and IELM, Meanwhile, the tradeoff parameter C for ELM and IELM was chosen from a set $\{2^{-8}, \cdots, 2^{8}\}$, and the number of hidden layer node is set to 500. For the fairness of this experiments, ELM and IELM use the same penalty parameter C (we tried to find penalty parameter C which make ELM and IELM all achieve better results).

4.3 Experimental results and analysis

4.3.1 Indian Pines remote sensing image data results

In this section we show the comparison of IELM with CRNN, MSMRF, SVM, ELM for classification. The classification results of the five classification methods with Indian Pines of

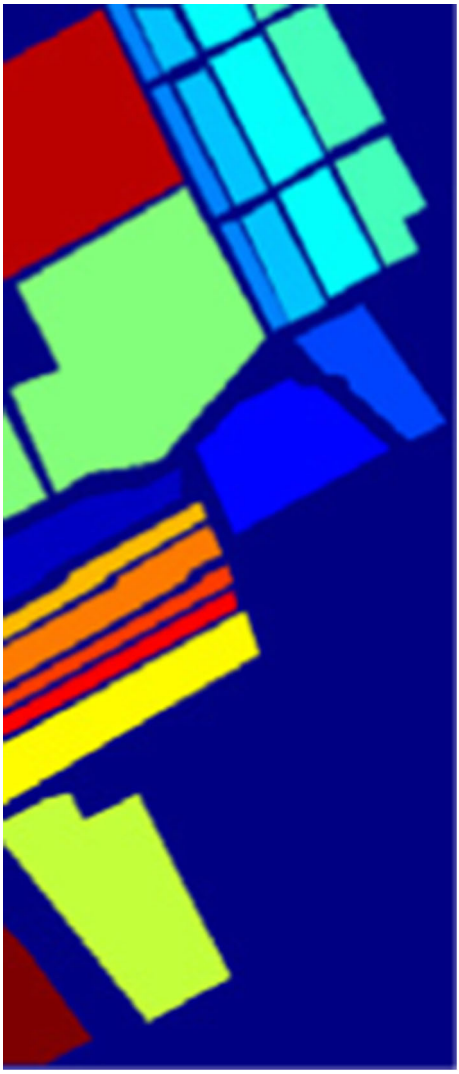

Fig. 2 Ground truth of Salinas scene

remote sensing image classification results on the data are shown in Table 4. The classification results of the four classification methods of Indian Pines of remote sensing image are shown in Fig. 3.

From Table 4, we can see that compared with SVM, CRNN, ELM algorithm, the precision, average accuracy, overall accuracy and Kappa coefficient the of IELM in each category have been enhanced. On category Alfalfa, the classification accuracy of SVM, ELM algorithms maintain at a relatively low level, IELM is much higher than that of the two algorithms. CRNN achieve good results, this is because it exploits the discriminative nature of Collaborative Representation for classification. Although on some categories of data the accuracy IELM algorithm is lower than the other algorithms, but average accuracy, overall accuracy and Kappa coefficient are superior to SVM, CRNN, ELM algorithm. IELM algorithm outperforms other algorithms because IELM takes geometric characteristics and discrimination information of

Table 3 Salinas scene category information data sets

Class	Samples
Brocoli_green_weeds_1	2009
Brocoli_green_weeds_2	3276
Fallow	1976
Fallow_rough_plow	1394
Fallow_smooth	2678
Stubble	3959
Celery	3579
Grapes_untrained	11,271
Soil_vinyard_develop	6203
Corn_senesced_green_weeds	3278
Lettuce_romaine_4wk	1068
Lettuce_romaine_5wk	1927
Lettuce_romaine_6wk	916
Lettuce_romaine_7wk	1070
Vinyard_untrained	7268
Vinyard_vertical_trellis	1807
Total	54,129

data into account to form a reasonable model. The experiments indicate: MSMRF is better than IELM on average accuracy, overall accuracy and Kappa coefficient, on some categories of data the accuracy IELM is better than that of MSMRF. The run time of IELM is longer than that of SVM and ELM, but shorter than that of CRNN and MSMRF.

Fig. 3 shows the classification diagram of the five classification algorithms with Indian Pines of remote sensing image. Compared with Fig.1 we can see on red color region, IELM and SVM have better classification results than that of CRNN and ELM. For misclassification

Table 4 Comparison of different classification algorithms in Indian Pines data classification (%)

Class	SVM	ELM	CRNN	MSMRF	IELM
Alfalfa	0.00	28.30	98.11	98.11	100.00
Corn-notill	74.91	59.62	79.42	80.13	74.14
Corn-mintill	38.30	52.24	51.64	57.33	57.33
Corn	0.87	47.62	47.19	51.95	43.72
Grass-pasture	66.26	68.90	72.76	71.75	70.12
Grass-trees	90.51	66.98	87.28	87.69	90.26
Grass-pasture-mowed	0.00	84.00	100.00	100.00	100.00
Hay-windrowed	97.93	51.03	78.72	84.71	100.00
Oats	0	36.84	94.74	94.74	89.74
Soybean-notill	39.04	76.62	73.90	77.77	70.15
Soybean-mintill	93.98	86.45	80.72	83.95	86.86
Soybean-clean	6.92	55.52	66.89	71.50	60.92
Wheat	93.78	75.60	96.65	98.09	100.00
Woods	92.82	57.38	94.69	96.64	95.55
Buildings-Grass-TreesDrives	86.97	53.72	94.68	95.74	89.10
Stone-Steel-Towers	0	20.21	68.09	85.11	70.12
Feature number	(75)	(75)	(75)	(30)	(75)
OA	70.93	66.39	78.62	81.41	80.34
AA	48.89	57.57	80.34	83.45	81.00
Kappa	65.63	60.57	75.69	78.84	77.23
Training time(seconds)	0.0312	0.0624	-	-	0.3588
Testing time(seconds)	0.1872	0.3276	-	-	8.6893
Total time(seconds)	0.2184	0.3900	39.5619	27.8618	9.0481

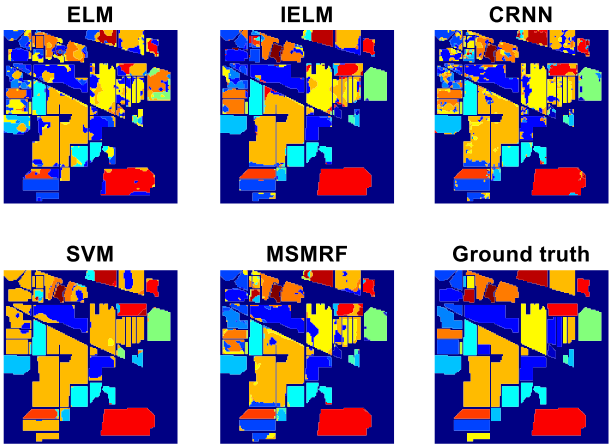

Fig. 3 Different classification algorithms classification results

phenomenon, SVM is most obviously, IELM is better than that of other three algorithms. This further confirms the effectiveness of IELM.

4.3.2 Salinas scene remote sensing image data results

To further demonstrate the effectiveness of the proposed algorithm, this section we have done experiments with Salinas scene of remote sensing image data to compared IELM with CRNN, SVM, MSMRF, ELM. The results of Salinas scene of remote sensing image classification on

Table 5 Different classification algorithms compare classification results with Salinas scene data (%)

Class	SVM	ELM	CRNN	MSMRF	IELM
Brocoli_green_weeds_1	100.00	100.00	99.80	100.00	100.00
Brocoli_green_weeds_2	100.00	99.89	99.46	100.00	99.86
Fallow	66.16	99.34	98.36	99.80	99.95
Fallow_rough_plow	96.30	93.84	82.03	97.46	98.70
Fallow_smooth	98.11	99.09	98.19	99.25	99.77
Stubble	100.00	100.00	99.64	100.00	100.00
Celery	100.00	100.00	99.92	100.00	98.76
Grapes_untrained	95.64	94.50	98.11	98.63	98.12
Soil_vinyard_develop	100.00	100.00	100.00	100.00	100.00
Corn_senesced_green_weeds	97.87	98.34	97.50	99.82	99.72
Lettuce_romaine_4wk	100.00	100.00	93.85	100.00	99.43
Lettuce_romaine_5wk	100.00	99.32	98.85	100.00	100.00
Lettuce_romaine_6wk	99.34	99.89	97.13	99.12	98.34
Lettuce_romaine_7wk	99.43	98.39	86.02	99.81	99.43
Vinyard_untrained	49.20	67.42	83.18	93.63	97.57
Vinyard_vertical_trellis	99.66	98.66	97.99	99.50	96.31
Feature number	(75)	(75)	(75)	(30)	(75)
OA	90.68	94.04	95.95	98.70	98.95
AA	93.86	96.79	95.63	99.19	99.12
Kappa	89.58	93.35	95.49	98.55	98.83
Training time(seconds)	0.0312	0.1248	-	-	0.4368
Testing time(seconds)	3.1200	1.2324	-	-	1.7160
Total time(seconds)	3.1512	1.3572	2384.4	3796.1	2.1528

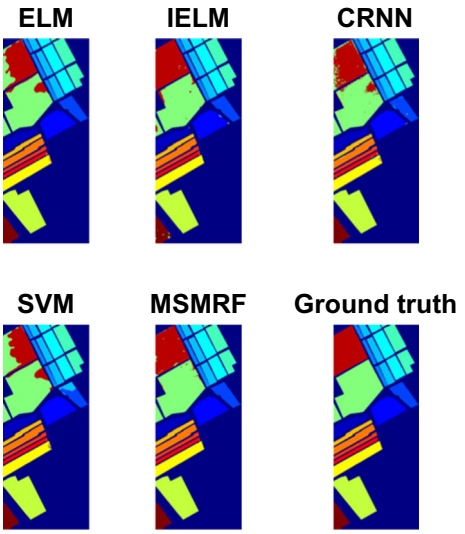

Fig. 4 Different classification algorithms classification results

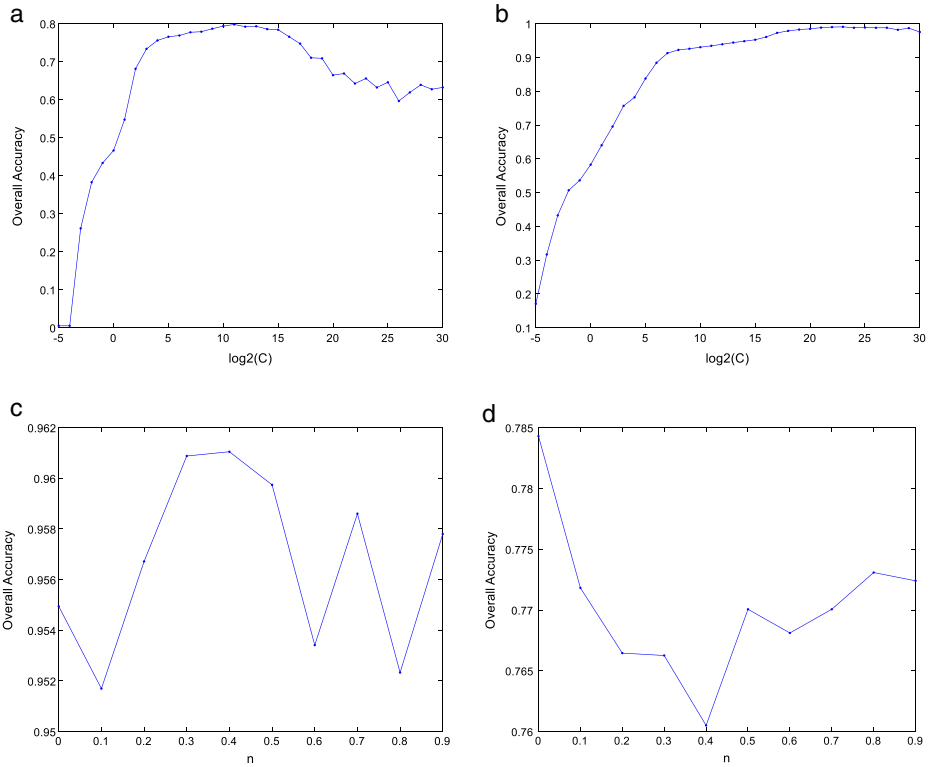

Fig. 5 Impacts on OA of IELM by different factors. (**a**) On Indian Pines data (penalty parameter *C*) (**b**) On Salinas scene data (penalty parameter *C*) (**c**) On Indian Pines data (parameter *n*) (**d**) On Salinas scene data (parameter *n*)

the data are shown in Table 5. The results of Salinas scene of remote sensing image classification are shown in Fig. 4.

Table 5 shows that SVM, MSMRF, ELM, CRNN, IELM have achieved good results, but IELM is more effective than that of other algorithms. Especially in category Vinyard_untrained data the classification accuracy of SVM, CRNN, ELM are maintain at a relatively low level, but the accuracy of IELM is much higher. Fig. 4 shows the classification diagram of the five algorithms. Comparing Fig. 4 with Fig. 2 we can see that for red color region, IELM is better than that of other three algorithms. With ELM and SVM there are large scale misclassification phenomenon.

4.4 Parameter analysis of IELM

In this section we analyze how the parameter n and C influence the performance of IELM. IELM uses Cosine as activation function, the number of hidden node is 500. The experiments results are shown in Fig. 5, in which different values of the parameter and corresponding classification accuracy are demonstrated.

From Fig. 5 we can see that different value of parameter n and C bring different result. For Indian Pines data, $n = 0.4$,C=$2^{10} \sim 2^{15}$ is best; for Salinas scene data $n = 0$, $C = 2^{20} \sim 2^{30}$is best. The performance of IELM can be adjusted by changing parameters.

4.5 Investigation on the computation cost

The run time of ELM,SVM,CRNN,MSMRF, IELM on Indian Pine and Salinas scene data sets are shown in Table 4 and Table 5. From Table 4 we can see: SVM being with shortest training and testing time, MSMRF and CRNN being with longest run time. Table 5 shows SVM being with shortest training time, but testing time is longer than that of ELM and IELM, CRNN and MSMRF being with longest run time. From the two table it can be concluded that on Indian Pines data SVM is with best computation efficiency, on Salinas scene data ELM and IELM have better efficiency than that of SVM, CRNN and MSMRF. It is worth noting that Salinas scene data is much large than Indian Pines data, this suggests that ELM and IELM is promising for being used with bigger data classification.

5 Conclusions

This paper proposes a hyperspectral remote sensing image classification algorithm based on information discriminative extreme learning machine. The innovation of proposed method is that it considers the connection and difference information of spectral remote sensing image data, the proposed algorithm IELM introduces the concept of the intra-class scatter and within class scatter, reflecting the discrimination information of input space data. By maximizing the intra-class scatter and minimizing within class scatter, it optimizes output weights of extreme learning machine. Comparison experiment results show that the classification effect of the proposed method is better than that of SVM, CRNN, ELM algorithm.

Acknowledgements The work was supported by National Natural Science Foundation of China (61105085) and Science Foundation of education ministry of Liaoning province (L2014427).

References

1. Baassou B, He M-Y, Mei S-h (2013) An accurate SVM-based classification approach for hyperspectral image classification, 21st International Conference on Geoinformatics
2. Bazi Y (2014) Differential evolution extreme learning machine for the classification of hyperspectral images. IEEE Geosci Remote Sens Lett 11(6):1066–1070
3. Bioucas-Dias J, Figueiredo MAT (2010) Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing, in Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process. Evol Remote Sens (WHISPERS) 1–4
4. Bruzzone L, Chi M, Marconcini M (2006) A novel transductive SVM for the semisupervised classification of remote sensing images. IEEE Trans Geosci Remote Sens 44(11):3363–3373
5. Camps-Valls G, Bruzzone L (2005) Kernel-based methods for hyperspectral image classification. IEEE Trans Geosci Remote Sens 43(6):1351–1362
6. Camps-Valls G, Gomez-Chova L, Muñoz-Marí J, Vila-Francés J (2006) Composite kernels for hyperspectral image classification. IEEE Geosci Remote Sens Lett 3(1):93–97
7. Di W, Crawford M-M (2011) Active learning via multi-view and local proximity co-regularization for hyperspectral image classification. IEEEJ Sel Topics Signal Process 5(3):618–628
8. Gualtieri JA, Cromp RF (1999) Support vector machines for hyperspectral remote sensing classification, in Proc. SPIE 27th AIPR workshop: Adv. Comput. Assisted Recognit Int Soc Opt Photonics, Washington, DC, USA, 221–232
9. Heras DB, Argüello F, Quesada-Barriuso P (2014) Exploring ELM-based spatial-spectral classification of hyperspectral images. Int J Remote Sens 35(2):401–423
10. Huang G-B (2014) An insight into extreme learning machines: random neurons, random features and kernels. Cogn Comput 6(3):376–390
11. Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B 42(2):513–529
12. Jun G, Ghosh J (2013) Semisupervised learning of hyperspectral data with unknown land-cover classes. IEEE Trans Geosci Remote Sens 51(1):273–282
13. Li H, Jiang T, Zhang K (2003) Efficient robust feature extraction by maximum margin criterion[C]. Adv Neural Inf Proces Syst 16
14. Li J, Bioucas-Dias J, Plaza A (2013) Semi-supervised hyperspectral image classification using soft sparse multinomial logistic regression. IEEE Geosci Remote Sens Lett 10(2):318–322
15. Li W, Du Q, Zhang F (2015) Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. IEEE Geosci Remote Sens Lett 12(2):389–393
16. Liang N-Y, Huang G-B, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17(6):1411–1423
17. Lu X-Q, Y Y, Zheng X-T (2016) Joint dictionary learning for multispectral change detection. IEEE Transactions on Cybernetics PP(99):1–14
18. Ma L, Crawford M-M, Tian J (2010) Local manifold learning based k-nearest-neighbor for hyperspectral image classification. IEEE Trans Geosci Remote Sens 48(11):4099–4109
19. Man Z-H, Wang D-H, Cao Z-W, Khoo SY (2012) Robust single-hidden layer feedforward network-based pattern classifier. IEEE Transactions on Neural Networks and Learning Systems 23(12):1974–1986
20. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans Geosci Remote Sens 42(8):1778–1790
21. Pal M (2009) Extreme-learning-machine-based land cover classification. Int J Remote Sens 30(14):3835–3841
22. Pal M, Maxwell AE, Warner TA (2013) Kernel-based extreme learning machine for remote-sensing image classification. Remote Sens Lett 4(9):853–862
23. Peng Y, Lu B-L (2015) Discriminative graph regularized extreme learning machine and its application to face recognition. Neurocomputing 149:340–353
24. Pu H, Chen Z, Wang B (2014) A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery. IEEE Trans Geosci Remote Sens 52(11):7008–7002
25. Rajan S, Ghosh J, Crawford MM (2008) An active learning approach to hyperspectral data classification. IEEE Trans Geosci Remote Sens 46(4):1231–1242
26. Rajesh R, Siva Prakash J (2011) Extreme learning machines - a review and state-of-the-art. International Journal of Wisdom based Computing 1(1):35–49

27. Rong H-J, Huang G-B, Sundararajan N, Saratchandran P (2009) Online sequential fuzzy extreme learning machine for function approximation and classification problems, IEEE transactions on systems, man, and cybernetics. Part B: Cybernetics 39(4):1067–1072
28. Rumelhart D, Hinton G, Williams R (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536
29. Samat A, Du P, Liu S, Li J, Cheng L (2014) E^2LMs: ensemble extreme learning machines for hyperspectral image classification. IEEE J Sel Topics Appl Earth Observ Remote Sens 7(4):1060–1069
30. Sami ul Haq Q, Tao L, Yang S (2011) Neural network based adaboosting approach for hyperspectral data classication. International Conference on Computer Science and Network Technology (ICCSNT), 241–245
31. Tang J, Deng C, Huang G-B (2015) Extreme learning machine for multilayer perceptron. IEEE Trans Neural Netw Learn Syst 27(4):809–821
32. Tao D-P, Jin L-W, Liu W-F, Li X-L (2013) Hessian regularized support vector Machines for Mobile Image Annotation on the cloud. IEEE Transactions on Multimedia 15(4):833–844
33. Tao D-P, Guo Y-N, Li Y-T (2016) Person Re-identification by dual-regularized KISS metric learning. IEEE Trans Image Process 25(6):2726–2738
34. Tarabalka Y, Benediktsson JA, Chanussot J (2009) Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques. IEEE Trans Geosci Remote Sens 47(8):2973–2987
35. Wang D-H, Huang G-B (2005) Protein sequence Classificati -on using extreme learning machine. Proceedings of International Joint Conference on Neural Networks 3:1406–1411
36. Wang Q, Lin J-Z, Yuan Y (2016) Salient band selection for hyperspectral image classification via manifold ranking. IEEE Transactions on Neural Networks and Learning Systems 27(6):1279–1289
37. Yu Q, Miche Y, Eirola E, van Heeswijk M, Severin E, Lendasse A (2013) Regularized extreme learning machine for regression with missing data. Neurocomputing 102:45–51
38. Yuan Y, Lin J-Z, Wang Q (2016) Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization. IEEE Transactions on Cybernetics 46(12):2966–2977
39. Zhao J-W, Wang Z-H, Park D-S (2012) Online sequential extreme learning machine with forgetting mechanism. Neurocomputing 87:79–89
40. Zhou Y-C, Wei Y-T (2016) Learning hierarchical spectral–spatial features for hyperspectral image classification. IEEE Transactions on Cyberentics 46(7):1667–1678
41. Zong W-W, Huang G-B, Chen Y (2013) Weighted extreme learning machine for imbalance learning. Neurocomputing 101:229–242

Deqin Yan Professor at College of computer and information technoloy, Liaoning Normal University. He received Ph.D. at Nankai University in 1999. His research interest is pattern recognition. Email: yandeqin@163.com

Yonghe Chu Master graduate student at College of computer and information technology, Liaoning Normal University. His research interest is information processing and pattern recognition. Email: 18842878491@163.com

Lina Li Master graduate student at College of computer and information technology, Liaoning Normal University. Her research interest is information processing and pattern recognition. Email: 454402989@qq.com

Deshan Liu Associate Professor at College of computer and information technology, Liaoning Normal University. His research interest is information processing. Email: deshanliu@yeah.net