

PERFORMANCE EVALUATION OF ELM WITH A-OPTIMIZED DESIGN REGULARIZATION FOR REMOTE SENSING IMAGERY CLASSIFICATION

Yi Lin¹, Tinghui Zhang¹, Kun Qian², Guangshun Xie¹, Jianqing Cai^{2*}

¹ School of Surveying and Geo-Informatics, Tongji University, (linyi, zhang_th)tongji.edu.cn, 786174501@qq.com

² Institute of Geodesy, University of Stuttgart, kun.qian@dlr.de, cai@gis.uni-stuttgart.de

Commission I, WG I/2

KEY WORDS: remote sensing, land use classification, extreme learning machine, regularization

ABSTRACT:

The automatic classification technology of remote sensing images is the key technology to extract the rich geo-information in remote sensing images and to monitor the dynamic changes of land use and ecological environment. Remote sensing images have the characteristics of large amount of information and many dimensions. Therefore, how to classify and extract the information in remote sensing images has become a crucial issue in the field of remote sensing science. With the development of neural network theory, many scholars have carried out research on the application of neural network models in remote sensing image classification. However, there are still some problems to be solved in artificial neural network methods. In this study, considering the problem of large-scale land classification for medium resolution and multi-spectral remote sensing imagery, an improved machine learning algorithm based on extreme learning machine for remote sensing classification has been developed via regularization theory. The improved algorithm is more suitable for the application of post-classification change monitoring of features in large scale imaging. In this study, our main job is to evaluate the performance of ELM with A-optimal design regularization (here we call it simply as A-optimal RELM). So the accuracy and efficiency of A-optimal RELM algorithm for remote sensing imagery classification, as well as the algorithms of support vector machine (SVM) and original ELM are compared in the experiments. The experimental results show that A-optimal RELM performs the best on all three different images with overall accuracy of 97.27% and 95.03% respectively. Besides, the A-optimal RELM performs better on the details of distinguish similar and confusing terrestrial object pixels.

1. INTRODUCTION

The remote sensing technology has been applied in many fields such as environment or urban monitoring. Multi-spectral remote sensing imagery has an enormous amount information of many types of objects on the earth. The information extraction from RS imagery is the basic requirement for the application and pro-analysis.

The automatic classification technology of remote sensing images is the key technology to extract the rich geo-information in remote sensing images and to monitor the dynamic changes of land use and ecological environment. Therefore, how to make full use of image information to identify and classify surface features is an eternal theme in the field of remote sensing technology efficiently and accurately.

For the application of large-scale land cover change detection based on post-classification, traditional machine learning method is more appropriate. Over the years, with the development of machine learning theory and technology, a variety of neural network methods have been proposed and utilized to solve image classification problems, including Radial Basis Function Neural Network (RBFNN), Multi-layer Perception (MLP)(Iversen et al., 2005), Self-organization Map Networks (SOM)(Giacco et al., 2010), Wavelet Neural Networks (WNN)(Angrisani et al., 2001) and other neural network classification methods. Various optimized algorithms have been developed based on those neural network methods to solve classification cases. Wavelet transform and WNN methods was used to detect and classify ephemeral signals simultaneously and automatically, using wavelet nodes to replace nodes in the first layer of the network. The experimental results are like those obtained by existing methods. MLP classifiers and autocorrelative neural networks (AANN) were combined to improve the classification discrimination between

feature space overlapping data effectively. Gao et al. combine SMOTE algorithm and Particle Swarm Optimization (PSO) to optimize the RBF model, and propose a classification method that can effectively handle the bicategorical imbalance problem(Gao et al., 2011). Feng et al. used dynamic BP algorithm to train the MLP learning model for mail classification study and showed a significant improvement in learning efficiency and classification accuracy compared to the traditional MLP model(Feng and Daqi, 2013).

These methods such as neuron network algorithms have gone through a process ranging from simple to complex, from specific to extensive and from single method to multi-combined.

However, there are still some problems to be solved in artificial neural network. For example, when dealing with a large amount of data in high-dimensional feature space, it is easy to fall into local minimum for nonlinear optimization problems, training efficiency is not high enough, a lot of parameters to be set manually, and the activation function should be differentiable. Especially when processing large amounts of data in high-dimensional feature space, the speed of ANN training is severely affected if the traditional methods of feedforward propagation and backward propagation are used to estimate the power array by random gradient descent(Abuelgasim et al., 1996). In order to train artificial neural networks, it often takes days or more. In addition, gradient descent-based algorithms converge easily to local minimum questions, resulting in low prediction accuracy relatively. Simultaneous iterative adjustment of parameters leads to dependencies between parameters. The algorithm also requires many iterative steps in order to obtain better generalization performance. In 2004, Huang G.B proposed the method of extreme learning machine (ELM) (Huang et al., 2004). The improved algorithm based on the ELM method also performs

well on large-scale remote sensing classification and can also be applied into the following change monitoring experiment and analysis (Lin et al., 2018). In this study, an optimal algorithm with A-optimized design regularization ELM is proposed and applied to real remote sensing classification experiments.

2. METHODS AND MATERIALS

2.1 Method

The extreme learning machine (ELM)(Huang et al., 2004) is a machine learning method of single-hidden layer of feed-forward networks (SLFNs), which assigns the input weights and number of hidden layers randomly. A simple learning method for SLFNs called extreme learning machine (ELM) can be summarized as follows:

Algorithm ELM. Given a training set $x = \{(x_i, y_i) | x_i \in \mathbf{R}^n, y_i \in \mathbf{R}^m, i = 1, \dots, N\}$, and activation function $\varphi(x)$, and the number of hidden nodes L

Step 1. Randomly assign input weight w_i and bias $b_i, i=1, \dots, L$.

Step 2. Calculate the hidden layer output matrix \mathbf{H} , which can be expressed as

$$\mathbf{H} = \begin{bmatrix} g(w_1x_1 + b_1) & \dots & g(w_Lx_1 + b_L) \\ \vdots & \dots & \vdots \\ g(w_1x_N + b_1) & \dots & g(w_Lx_N + b_L) \end{bmatrix}_{N \times L} \quad (1)$$

Step 3. Calculate the output weight $\beta: \beta = \mathbf{H}^\dagger \mathbf{Y}$, where $\mathbf{Y} = [y_1, \dots, y_N]^T$

where \mathbf{H} is feature mapping matrix, or hidden layer matrix of the neural network. \mathbf{Y} is output vector or matrix.

And \mathbf{H}^\dagger is the *Moore–Penrose generalized inverse* of matrix \mathbf{H} (Huang et al., 2011), which is generated by

$$\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \quad (2)$$

In practical application cases, there is still an important unsolved problem of ELM, which is how to obtain the most appropriate architecture of SLFNs. And as for the original ELM, there are some problems such as overfitting, non-optimal and reasonableness problem. The general method to improve the generalization ability and avoid over-fitting is regularization. It is important to select a proper regularization parameter for ELM. So, it is important to select a proper regularization parameter for the ELM. However, there is no general method to choose a proper or an optimal regularization parameter so far. Deng et al. have proposed a numerical heuristic method based on cross validation to determine the regularization parameter(Deng et al., 2009). The output function of ELM classifier is improved by the regularization parameter λ :

$$f(x) = h(x)\beta = h(x) \left(\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I} \right)^{-1} \mathbf{H}^T \mathbf{Y} \quad (3)$$

In such equation, we have the output \mathbf{Y} as a $N \times m$ matrix, when there are N distinct training samples and m output neurons. the output weight is thus a $L \times m$ matrix, where L represents the number of hidden neurons.

In order to acquire an optimal regularization parameter, we apply A-optimal design regularization, proposed by Cai (2004), to determinate the regularization parameter. To improve generalization performance and stability of ELM, in some applications regularization is introduced to penalize the weight matrix $\hat{\mathbf{B}}$. But how to obtain appropriate regularization

parameters is still a crucial issue worth discussed. In order to acquire an optimal regularization parameter, an A-optimal design regularization is introduced, which is realized by calculating the minimum trace of mean square error (MSE) of $\hat{\mathbf{B}}$ in this study. The A-optimal regularization is applied to determine the parameter. The regularization parameter λ follows by A-optimal designed regularization in the sense of minimizing the trace of MSE matrix:

$$\lambda = \frac{\text{trace} \left(\mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-3} \right)}{\mathbf{B}^T (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-2} \mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{I})^{-1} \mathbf{B}} \quad (4)$$

Finally, the solution of $\hat{\mathbf{B}}$ can be written into the following equations:

$$\begin{aligned} \beta_1 &= (\mathbf{H}^T \mathbf{H} + \lambda_1 \mathbf{I})^{-1} \mathbf{H}^T y_1 \\ \hat{\mathbf{B}} = \beta_2 &= (\mathbf{H}^T \mathbf{H} + \lambda_2 \mathbf{I})^{-1} \mathbf{H}^T y_2 \\ &\dots \\ \beta_m &= (\mathbf{H}^T \mathbf{H} + \lambda_m \mathbf{I})^{-1} \mathbf{H}^T y_m \end{aligned} \quad (5)$$

2.2 Experiment

To examine the practicability, two accuracy comparison experiments were conducted by real Landsat remote sensing imagery of different region. For the accuracy comparison experiment, two Landsat 8 OLI images of different terrain features and region area are utilized. The research area of the two experiments are shown in Figure 1, which are Wuhan East Lake in middle part of China, and Hamburg City in north Germany. In the research area of Wuhan East Lake, the land scape mainly consists of buildings, bare land, forest land, paddy and lakes of different shapes and size. In the study area of Hamburg, the land cover types are divided into bare land, building, paddy, forest, cement land and water.

The pre-processing includes radiometric calibration, atmosphere correction and feature space construction. The A-optimal RELM method proposed in this study is compared with other two methods: support vector machine (SVM) and standard ELM method.

In the pre-processing steps, the radiometric calibration and atmosphere correction were processed on all RS images in ENVI 5.1 to remove the influence of atmosphere radiomutation and to achieve real surface reflectance. For the feature space construction, seven spectral band was directly used as feature space in Wuhan imagery, while in the Hamburg imagery the principle component analysis (PCA) band and different index bands are adopted to build the feature space. For the sample dataset construction, the Google Earth high-resolution imagery is taken as a reference to pick up the samples. And for the input data in the experiment, the training and testing samples are generated from the dataset randomly. The main steps of this study are shown in Figure 2.



(a) Wuhan East Lake



(b) Hamburg

Figure 1. Research Area presented by Landsat 8 OLI RGB imagery

Region	Time	Sensors	Size
Hamburg	29/08/2018	Landsat 8	705×725
Wuhan	28/05/2014	OLI	598×597

Table 1. Landsat image information

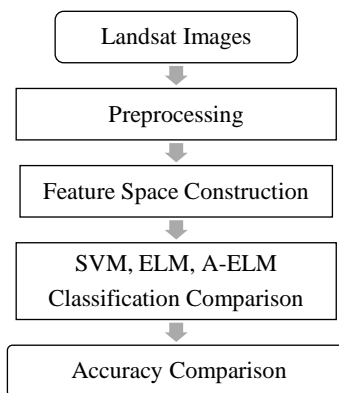


Figure 2. Main experimental procedure

3. RESULTS

To examine the accuracy of the A-optimal RELM method, a comparison experiment with other two methods: SVM and original ELM was generated by two Landsat8 OLI images of Wuhan East Lake and Hamburg. The classification results of Wuhan East Lake and Hamburg Landsat images are presented in Figure 3 and Figure 4 separately. To demonstrate the difference

more directly, the confusion matrix of SVM and A-ELM is also calculated in this experiment.

The kappa coefficient calculated by the following equations:

$$p_0 = \frac{\sum x_{kk}}{N}$$

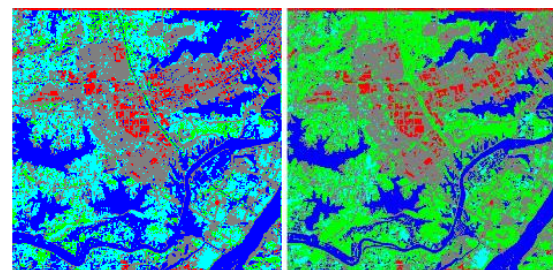
$$p_e = \frac{\sum_k n_{validation} \times n_{classification}}{N^2} \quad (6)$$

$$kappa = \frac{p_0 - p_e}{1 - p_e}$$

And N represents the number of sample pixels in each dataset.

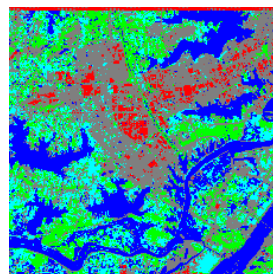
3.1 Classification comparison of Wuhan East Lake

In the classification result of standard ELM, most paddy pixels are misclassified into forest. And the boundary between watershed and land, especially in area with complex terrain, is not clear enough. The accuracy statistics of the three methods are shown in Table2. From the precision statistics table of Wuhan image, the A-optimal RELM method reaches the best result with the overall accuracy of 97.27% and the kappa coefficient of 0.9504, meanwhile the standard ELM has the lowest overall accuracy of 84.11%. It is obvious that most pixels of paddy are misclassified as forest by standard ELM in Figure 3 (b). The confusion matrix of the three methods are demonstrated in Table 3-5 to give a direct comparison. In the confusion matrix of SVM, it is obvious that a large amount of water pixels is misclassified to building because of the similar reflection feature. And through the result of A-optimal RELM, we can see that the phenomenon of misclassifying has been corrected effectively.



(a)SVM

(b) ELM



(c) A-optimal RELM

- Water
- Bareland
- Forest
- Paddy
- Building

Figure 3. classification results based on Wuhan East Lake image

Method	Accuracy	
	Overall precision	Kappa coefficient
ELM	84.11%	0.7596
SVM	92.64%	0.9069
A-ELM	97.27%	0.9504

Table 2. Accuracy comparison on Wuhan image experiment

SVM	Validation pixels				
	Water	Bare	Forest	Paddy	Building
water	730	0	0	0	0
Bare	21	952	13	2	1
Forest	0	6	503	3	0
Paddy	4	0	8	557	0
Building	229	8	0	0	513

Table 3. Confusion matrix of SVM on Wuhan image experiment

ELM	Validation pixels				
	Water	Bare	Forest	Paddy	Building
water	730	0	0	0	0
Bare	80	890	1	7	11
Forest	32	2	156	323	0
Paddy	0	0	0	569	0
Building	31	7	0	0	712

Table 4. Confusion matrix of ELM on Wuhan image experiment

A-ELM	Validation pixels				
	Water	Bare	Forest	Paddy	Building
water	720	0	0	0	0
Bare	2	946	31	3	7
Forest	0	2	497	12	1
Paddy	0	0	4	565	0
Building	2	18	3	12	715

Table 5. Confusion matrix of A-optimal RELM on Wuhan image experiment

3.2 Classification comparison of Hamburg

From the precision statistics of Hamburg image in Table 6, the A-optimal RELM method also performs best with the overall accuracy of 95.03% and the kappa coefficient of 0.9399. And SVM has the lowest overall accuracy of 90.51%. The confusion matrix of the three methods are demonstrated in Table 7-9. In the confusion matrix of SVM, misclassification occurs in several different landcover types. For example, part of bare land pixels is misclassified to building. The pixels of cement floor and building also misclassified to each other because of the similar spectral feature. And in the result of A-optimal RELM, the number of misclassification pixels is significantly reduced. Especially, the number of misclassification pixels from bare land to building has been reduced from 114 to 19. In the classification results in Hamburg, it is obvious that a fraction of water was misclassified to cement land by ELM in Figure 5. And the classification results of other land cover pixels are similar in Hamburg imagery.

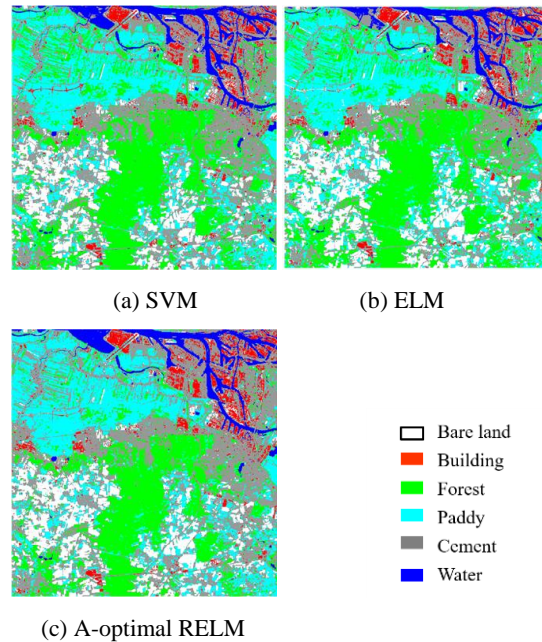


Figure 4. classification results based on Hamburg image

Method	Accuracy	
	Overall precision	Kappa
ELM	92.92%	0.8921
SVM	90.51%	0.8854
A-ELM	95.03%	0.9399

Table 6. Accuracy comparison on Hamburg image experiment

SVM	Validation pixels					
	Bare	Buildings	Forest	Paddy	Cement	Water
Bareland	525	8	1	0	1	0
Building	114	495	1	0	33	12
Forest	2	1	417	9	8	0
Paddy	0	0	6	316	3	0
Cement	16	29	15	6	432	2
Water	1	0	7	0	2	433

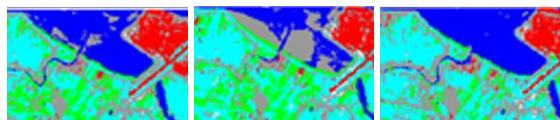
Table 7. Confusion matrix of SVM on Hamburg image experiment

ELM	Validation pixels					
	Bare	Buildings	Forest	Paddy	Cement	Water
Bareland	527	4	1	0	3	0
Building	26	543	0	3	72	11
Forest	1	0	417	14	5	0
Paddy	0	0	1	323	1	0
Cement	3	21	15	18	441	2
Water	0	1	2	0	1	439

Table 8. Confusion matrix of ELM on Hamburg image experiment

A-ELM	Validation pixels					
	Bare	Buildin gs	Forest	Paddy	Cement	Water
Bareland	531	3	1	0	0	0
Building	19	581	1	0	43	11
Forest	1	0	419	11	6	0
Paddy	1	0	0	324	0	0
Cement	3	14	11	15	456	1
Water	0	0	2	0	1	440

Table 9. Confusion matrix of A-optimal RELM on Hamburg image experiment



(a) SVM (b) ELM (c) A-optimal RELM

Figure 5. Part of misclassification details of watershed and cement land in Hamburg image

The experimental results show that A-optimal RELM performs the best on two different images with overall accuracy of 97.27% and 95.03% respectively. Overall, compared with SVM and ELM, A-optimal RELM can reach the highest precision and it is also appropriate for multi-spectral RS imagery classification. Besides, the A-optimal RELM performs better on the details of distinguish similar and confusing terrestrial object pixels. It indicates that based on the classification experiment from the high-precision A-optimal RELM algorithm, convincing followed-up analysis can also be carried out through the accurate classification results.

4. CONCLUSIONS

In this research, the results of the two comparison experiments demonstrate that the A-optimal ELM has a higher classification accuracy than SVM and original ELM methods. And the classification results on different Landsat remote sensing images have shown a better stability.

In conclusion, the A-optimal ELM is more stable and effective than a standard ELM. On the other hand, the experimental results demonstrate that the A-ELM has a higher classification accuracy than the other two classification methods. Additionally, based on the classification results, a simple spatial-temporal analysis about the LUCC in the Chaohu Lake basin is carried out. It can directly show the change status in the region spatially and temporally. Overall, compared with SVM and ELM, A-optimal RELM can reach the highest precision and it is appropriate for multi-spectral RS imagery classification. Besides, the A-optimal RELM performs better on the details of distinguish similar and confusing terrestrial object pixels. It indicates that utilizing the high-precision A-optimal RELM algorithm, convincing followed-up analysis can also be carried out based on the accurate classification results. For the future work, performance on other remote sensing imagery with different feature space construction need to be conducted: high-resolution images such as GF satellite, World View-2 or Quickbird, etc.

ACKNOWLEDGEMENTS

Sincere thanks are expressed to the reviewers for their insightful remarks which will improved the quality of this paper. This work was supported by the National Natural Science Foundation of China under Grant 41771449, the National Key Research and Development Program of China under Grant 2018YFB0505400, the DAAD Thematic Network Project No. 57421148, and the

International Exchange Program for Graduate Students, Tongji University.

REFERENCES

- Abuelgasim, A.A., Gopal, S., Irons, J.R. and Strahler, A.H., 1996. Classification of ASAS multiangle and multispectral measurements using artificial neural networks. *Remote Sensing of Environment*, 57(2): 79-87.
- Angrisani, L., Daponte, P. and Dapuzzo, M., 2001. Wavelet network-based detection and classification of transients. *IEEE Transactions on Instrumentation and Measurement*, 50(5): 1425-1435.
- Deng, W., Zheng, Q. and Chen, L., 2009. Regularized Extreme Learning Machine, computational intelligence and data mining, pp. 389-395.
- Feng, Q. and Daqi, G., 2013. Dynamic learning algorithm of multi-layer perceptrons for letter recognition, *international joint conference on neural network*, pp. 1-6.
- Gao, M., Hong, X., Chen, S. and Harris, C.J., 2011. A combined SMOTE and PSO based RBF classifier for two-class imbalanced problems. *Neurocomputing*, 74(17): 3456-3466.
- Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S. and Marinaro, M., 2010. Uncertainty Analysis for the Classification of Multispectral Satellite Images Using SVMs and SOMs. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10): 3769-3779.
- Huang, G., Zhou, H., Ding, X. and Zhang, R., 2011. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2): 513-529.
- Huang, G., Zhu, Q. and Siew, C., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks, *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*. IEEE, pp. 985-990.
- Iversen, A., Taylor, N.K. and Brown, K., 2005. Classification and verification through the combination of the multi-layer perceptron and auto-association neural networks, *international joint conference on neural network*, pp. 1166-1171.
- Lin, Y., Yu, J., Cai, J., Sneeuw, N. and Li, F., 2018. Spatio-temporal analysis of wetland changes using a kernel extreme learning machine approach. *Remote Sensing*, 10(7): 1129.