

Classification of Hyperspectral Images of Small Samples Based on Support Vector Machine and Back Propagation Neural Network

Cheng-Biao Fu and An-Hong Tian*

College of Information Engineering, Qujing Normal University,
Qujing, Yunnan 655011, China

(Received February 2, 2019; accepted December 10, 2019)

Keywords: hyperspectral image, classification accuracy, support vector machines, BP neural network

High-precision hyperspectral image classification when the number of samples is small is the focus of research in the field of hyperspectrum. At present, there are few studies on the effect of different training set samples on classification accuracy. To determine the effect of different training set samples on the classification accuracy of a hyperspectral image, the hyperspectral image of an Indian Pines farm is used as the data source. In this work, we study the classification accuracy results of support vector machine (SVM) and back propagation (BP) neural network when the training set samples are 1, 2, 5, 10, and 20%. Simulation results show that the overall accuracy (OA), average accuracy (AA), and Kappa coefficients of SVM and BP increase continuously with the number of samples in the training set. Under different numbers of training set samples, the classification accuracy of BP is greater than that of SVM. When the number of samples in the training set is 20%, the recognition accuracy of the BP classification method for seven features (Grass-pasture, Grass-trees, Hay-windrowed, Oats, Wheat, Woods, and Stone-Steel-Towers) is higher than 90%, and the recognition accuracy of Hay-windrowed features is 93.97%.

1. Introduction

Hyperspectral imagers can acquire spectral imaging data from hundreds of narrow and continuous bands in the visible-to-infrared spectral region. Each pixel of a hyperspectral image exists in the form of a vector, whose different elements correspond to spectral response values at different wavelength bands. Since different substances reflect different electromagnetic energies on a specific wavelength band, different substances can be distinguished on the basis of their spectral characteristics. Hyperspectral images have the characteristics of high spectral resolution, wide spectral range, and strong spectral correlation.^(1–3) They can detect the feature category that cannot be detected by multispectral images and are widely used in environmental detection, military security, astronomy, forestry protection, mineralogy, and other fields.

Classifying each pixel in a hyperspectral image is an issue that should be resolved in an application. According to the current classification method, the hyperspectral image

*Corresponding author: e-mail: tianah@mail.qjnu.edu.cn
<https://doi.org/10.18494/SAM.2020.2669>

classification method based on machine learning is the mainstream method of feature classification, which mainly uses a pattern recognition classifier to classify ground objects. For example, Zhang *et al.* determined the spatial information in the homogeneous region using the relative homogeneity coefficient through the Markov random field and combined the spectral features with the spatial information to effectively improve the classification effect of the support vector machine (SVM) classifier.⁽⁴⁾ Chen *et al.* used the window method to introduce spatial information through the joint sparse representation of pixels in the window, in order to minimize the reconstruction error and obtain the classification result.⁽⁵⁾ Guo *et al.* adopted the support tensor machine (STM) method to solve the hyperspectral classification problem and introduced the spatial information through the special structure of tensor feature expression to realize the hyperspectral classification of a spectral-spatial information joint.⁽⁶⁾ Gurram and Kwon proposed a contextual SVM classification method based on kernel space embedding by introducing spatial information by averaging it in the nuclear Hilbert space.⁽⁷⁾

However, the training set of small samples is one of the key points in the research on hyperspectral image classification, but the effect of different training set samples on classification accuracy has been rarely studied. The SVM maps the original feature vector to the high-dimensional space through the kernel function and realizes the classification by establishing the decision surface, which has strong small sample training classification ability. The back propagation (BP) neural network has a high classification accuracy, a strong parallel processing ability, a strong robustness and fault tolerance to noise, and can fully approximate complex nonlinear relationships. Therefore, to determine the effect of different training set samples on the classification accuracy of a hyperspectral image, the hyperspectral image of an Indian Pines farm is used as the data source. In this paper, when the training set samples are 1, 2, 5, 10, and 20%, the classification accuracies of SVM and BP are compared and analyzed.

2. Materials and Methods

2.1 Data source

The Indian Pines experimental dataset is an image of a farmland in Indiana, USA, taken in 1992 with the AVIRIS sensor. The image resolution is 145×145 pixel, the spatial resolution is 20 m, the spectral band range is 400–2500 nm, and the spectral resolution is 9.7–12 nm. After removing the bands with severe atmospheric absorption and noise effects, the remaining 200 bands were used for experiments. The data sample of the image is 10249 in total and contains 16 types of feature information. The real object image is shown in Fig. 1. The specific category name and the number of pixels contained in each class are shown in Table 1.

2.2 SVM and BP classifier

SVM is a typical supervised classification model.⁽⁸⁾ The basic principle is to find the classification hyperplane so that the two sample points can be separated. In fact, it is a problem to solve a convex optimization. Let $x_i \in R^d$ denote the training sample and $y_i \in \{+1, -1\}$ denote the corresponding category label, then the optimization problem of SVM can be described as

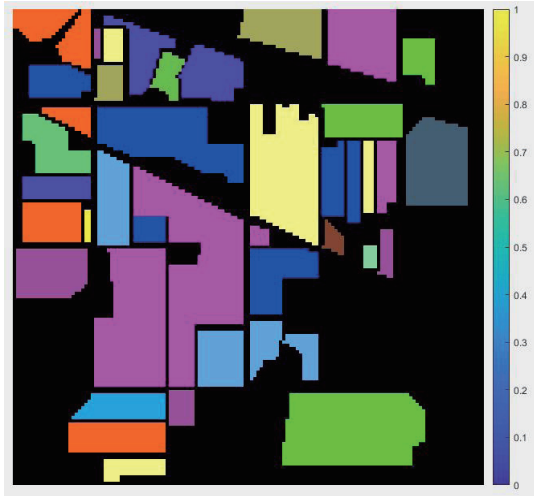


Fig. 1. (Color online) Real object image.

Table 1

Category names and numbers of pixels of 16 features.

Number	Class	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

$$\min \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad (1)$$

s. t.

$$y_i [(w \cdot \Phi(x_i)) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n,$$

where w is the weight vector, b is the intercept of the decision function, ζ_i is the slack variable, C is the penalty coefficient, and $\Phi(\cdot)$ is the nonlinear mapping function.

The BP neural network is a model for dealing with nonlinear problems.⁽⁹⁾ It is generally composed of input, hidden, and output layers. The basic principle is to feed back the results of neural network learning to the hidden layer and adjust the weights and thresholds so that the total error is minimized to meet the expected learning requirements. The specific calculation steps of BP are as follows:

Step 1: Calculate the middle layer output.

$$x_j^k = \frac{1}{1 + \exp(1 - \sum_{i=1}^m W_{ij} x_i^k + \theta_j)}, \quad j = 1, 2, \dots, q, \quad (2)$$

where q is the number of middle layers, m is the number of input layers, k is the number of learning, W_{ij} is the connection weight from the input layer to the middle layer, and θ_j is the threshold of the middle layer.

Step 2: Calculate the output layer output.

$$y_t^k = \frac{1}{1 + \exp(1 - \sum_{j=1}^n V_{jt} z_j^k + v_t)}, \quad t = 1, 2, \dots, n, \quad (3)$$

where n is the number of output layers, V_{ji} is the connection weight of the middle layer to the output layer, and v_i is the output layer threshold.

Step 3: Reverse transmission, calculate the error between the weight, the threshold, and the set value, and continuously update the weight and threshold in the neural network, so that the output of the output layer is made to be as close as possible to the expected output.

2.3 Classification accuracy evaluation index

After the hyperspectral images are classified, it is necessary to evaluate objectively the classification results. Generally, four indicators are often used for the evaluation, namely, class accuracy (CA), overall accuracy (OA), average accuracy (AA), and Kappa coefficient. CA is used to indicate the proportion of the correct classification result of each feature in the corresponding sample. OA indicates the proportion of the total correct classification results in all samples. AA represents the average of the proportions in which each category is correctly classified, taking into account the classification of each category. The Kappa coefficient is a ratio that takes into account the effect of uncertainty on the classification results and is used for consistency testing. The calculation formulas of these four evaluation indicators are as shown in Eqs. (4)–(7), respectively.

$$CA = R_{ii} / \sum_{j=1}^k R_{ij} \quad (4)$$

$$OA = \sum_{i=1}^k R_{ii} / T \quad (5)$$

$$AA = \left(\sum_{i=1}^k (R_{ii} / \sum_{i=1}^k R_{ij}) \right) / k \quad (6)$$

$$Kappa = (T(\sum_{i=1}^k R_{ii}) - \sum_{i=1}^k (\sum_{j=1}^k R_{ij} \sum_{j=1}^k R_{ji})) / (T^2 - \sum_{i=1}^k (\sum_{j=1}^k R_{ij} \sum_{j=1}^k R_{ji})) \quad (7)$$

Here, in practical applications, it is assumed that there are k categories of features, and by comparing the classification results with the real results, the confusion matrix \mathbf{R} of $k \times k$ can be obtained. \mathbf{R} represents the number of samples in which category j is recognized as i . T denotes the total number of the test samples.

3. Results and Discussion

In each ground, 1, 2, 5, 10, and 20% are selected as training set samples, and the remaining samples are used as test set samples to verify the classification accuracies of SVM and BP classifiers in different training set ratios. This is shown in Table 2. It can be seen from Table 2

Table 2

Classification accuracies of SVM and BP in different training set ratios (unit: %).

Classification	1% ratio		2% ratio		5% ratio		10% ratio		20% ratio	
	SVM	BP	SVM	BP	SVM	BP	SVM	BP	SVM	BP
Alfalfa	44.12	31.82	72.00	66.67	29.17	35.29	80.00	76.19	58.70	58.97
Corn-notill	45.21	53.07	56.06	67.81	67.05	74.38	76.98	80.94	82.51	84.14
Corn-mintill	58.10	53.05	49.80	56.11	66.44	60.59	78.34	78.13	80.55	78.74
Corn	41.57	18.13	38.99	39.29	39.35	42.65	60.00	56.27	62.71	67.34
Grass-pasture	80.24	83.76	78.97	71.72	89.33	90.62	82.23	82.25	91.83	93.77
Grass-trees	74.45	73.44	75.84	77.57	85.73	87.47	90.30	92.55	89.56	92.36
Grass-pasture-mowed	0.00	0.00	90.48	40.91	88.89	42.86	84.00	88.00	90.48	62.96
Hay-windrowed	80.30	76.90	94.05	88.98	91.08	87.23	92.42	91.01	96.86	93.97
Oats	19.35	11.86	30.43	44.44	50.00	66.67	54.55	100.00	75.00	90.91
Soybean-notill	37.04	40.46	66.85	67.61	62.28	75.47	72.23	79.15	79.82	83.40
Soybean-mintill	59.70	58.27	64.42	65.51	73.48	74.44	80.53	83.04	81.46	84.48
Soybean-clean	49.04	43.64	55.35	50.10	79.30	64.26	83.65	78.13	88.28	85.52
Wheat	83.26	90.10	80.75	84.42	91.61	92.50	89.60	91.75	96.99	92.00
Woods	81.91	88.08	90.83	88.92	91.88	91.44	91.37	92.06	92.58	93.42
Buildings-Grass-Trees-Drives	37.18	43.95	46.06	61.29	53.44	68.00	77.09	74.73	75.00	81.12
Stone-Steel-Towers	98.28	23.08	100.00	87.78	100.00	96.20	98.63	97.10	95.83	91.43
OA	60.16	60.41	67.95	70.29	75.17	76.92	81.76	83.13	84.72	86.05
AA	55.61	49.35	68.18	66.19	72.44	71.88	80.75	83.83	83.63	83.41
Kappa	54.06	54.20	63.07	65.81	71.60	73.64	79.14	80.76	82.51	84.06

that as the number of samples in the training set increases, the OA, AA, and Kappa coefficients of SVM and BP continue to increase. This is because the more samples in the training set, the richer the mark information and the local structure information contained, and the selected features can save more global similarity or local geometric property information, so the classification accuracy is higher. Moreover, the classification accuracy of BP is higher than that of SVM under different numbers of training set samples.

When the number of samples in the training set is 1%, the OA accuracy of both SVM and BP is about 60%. When the number of samples in the training set is 2%, the OA accuracy of SVM is about 67% and that of BP is about 70%. When the number of samples in the training set is 5%, the OA accuracy of both SVM and BP is about 75%. When the number of samples in the training set reaches 10 and 20%, the OA accuracy of both SVM and BP is more than 80%.

In addition, when the number of training set samples is 20%, the BP classifier's recognition accuracy for seven objects (Grass-pasture, Grass-trees, Hay-windrowed, Oats, Wheat, Woods, and Stone-Steel-Towers) is higher than 90% and the recognition accuracy of the Hay-windrowed object is 93.97%. The recognition accuracy of the Soybean-clean object is about 85%. The recognition accuracies of the four objects (Corn-notill, Soybean-notill, Soybean-mintill, and Buildings-Grass-Trees-Drives) are between 80 and 85%. The recognition accuracy of the

Corn-mintill object is 78.74%. The recognition accuracy of the Corn object is 67.34%. The recognition accuracy of the Alfalfa object is the lowest (58.97%).

Figure 2 shows classification identification maps of SVM and BP in different training set ratios. When the number of samples in the training set is very small, the phenomenon of “pepper salt” in the classification maps of SVM and BP is very clear. When the number of samples in the training set is increasing, the phenomenon of “pepper salt” in the classification maps of SVM and BP is gradually decreasing. Moreover, the BP classification map is smoother than the SVM classification map under different training set ratios.

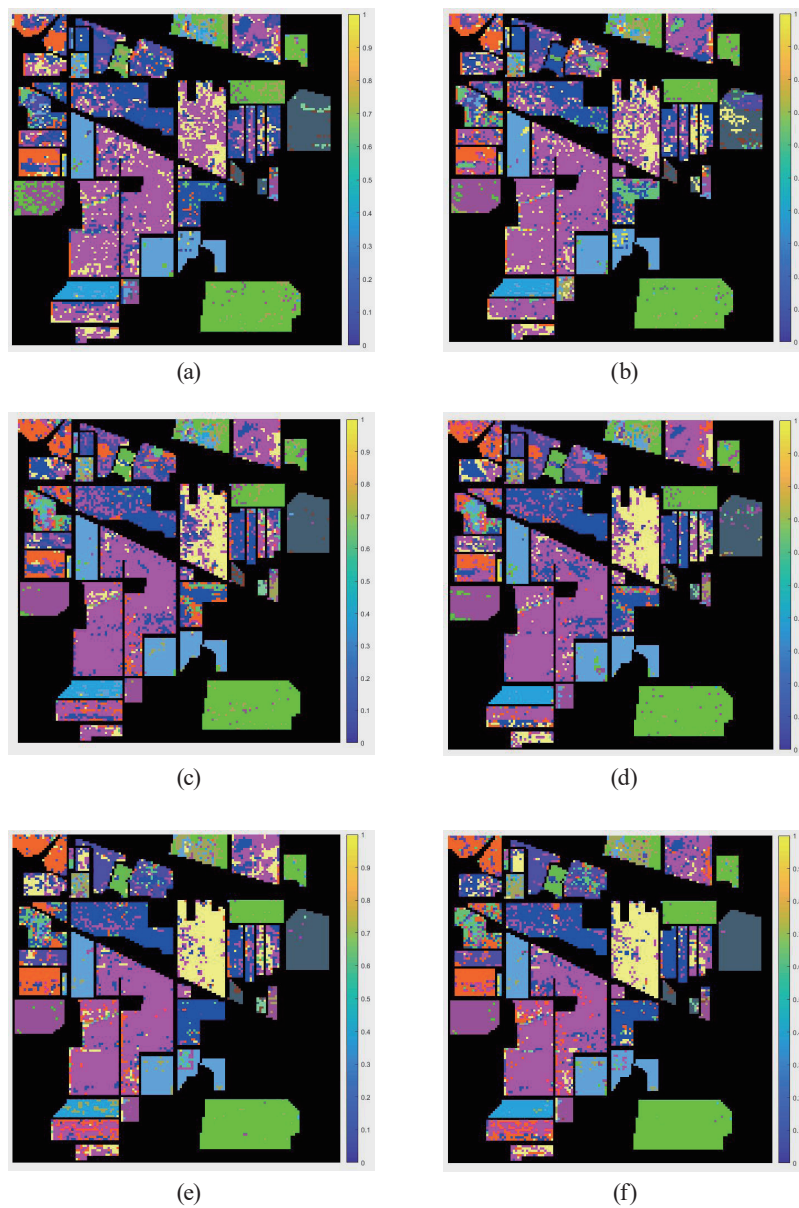


Fig. 2. (Color online) Classification identification maps of SVM and BP: (a) 1% of SVM, (b) 1% of BP, (c) 2% of SVM, (d) 2% of BP, (e) 5% of SVM, and (f) 5% of BP.

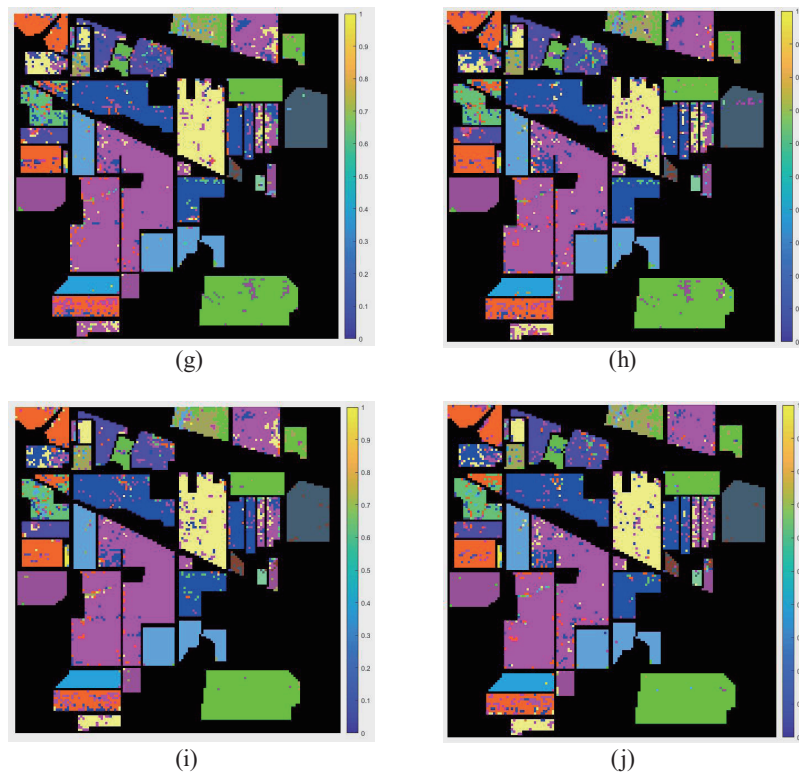


Fig. 2. (Color online) (Continued) Classification identification maps of SVM and BP: (g) 10% of SVM, (h) 10% of BP, (i) 20% of SVM, and (j) 20% of BP.

4. Conclusions

In this paper, the Indian Pines data set is used as the experimental data of a hyperspectral image, and 1, 2, 5, 10, and 20% are selected as the training set samples in 16 types of ground, and the remaining samples are used as test samples. The simulation results reveal the classification accuracies of SVM and BP classifiers in different training set ratios, which provide a reference for finding the best training set samples.

Acknowledgments

This work was financially supported by the National Natural Science Foundation of China under Grant No. 41901065.

References

- 1 M. A. Toksöz and İ. Ulusoy: IEEE Trans. Geosci. Remote Sens. **54** (2016) 4039. <https://doi.org/10.1109/TGRS.2016.2535458>
- 2 C. J. Bo, H. C. Lu, and D. Wang: IEEE Geosci. Remote Sens. Lett. **13** (2016) 177. <https://doi.org/10.1109/LGRS.2015.2504449>

- 3 E. L. Zhang, L. C. Jiao, X. G. Zhang, H. Y. Liu, and S. Wang: IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **9** (2016) 4160. <https://doi.org/10.1109/JSTARS.2016.2522182>
- 4 B. Zhang, S. S. Li, X. P. Jia, L. R. Gao, and M. Peng: IEEE Geosci. Remote Sens. Lett. **8** (2011) 973. <https://doi.org/10.1109/LGRS.2011.2145353>
- 5 Y. Chen, N. M. Nasrabadi, and T. D. Tran: IEEE Trans. Geosci. Remote Sens. **49** (2011) 3973. <https://doi.org/10.1109/TGRS.2011.2129595>
- 6 X. Guo, X. Huang, L. F. Zhang, L. P. Zhang, A. Plaza, and J. A. Benediktsson: IEEE Trans. Geosci. Remote Sens. **54** (2016) 3248. <https://doi.org/10.1109/TGRS.2016.2514404>
- 7 P. Gurram and H. Kwon: IEEE Geosci. Remote Sens. Lett. **10** (2013) 1031. <https://doi.org/10.1109/LGRS.2012.2227934>
- 8 J. López, S. Maldonado, and M. Carrasco: Neurocomputing **364** (2019) 227. <https://doi.org/10.1016/j.neucom.2019.07.072>
- 9 Y. An, X. C. Wang, Z. G. Qu, T. Liao, and Z. L. Nan: Optik **172** (2018) 753. <https://doi.org/10.1016/j.ijleo.2018.07.064>

About the Authors



Cheng-Biao Fu received his B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2005 and 2009, respectively. Since 2015, he has been a lecturer at Qijing Normal University. His research interests are in information identification and processing, hyperspectral remote sensing, and Internet of Things. (fucb@mail.qjnu.edu.cn)



An-Hong Tian received her B.S. and M.S. degrees from Chongqing University of Posts and Telecommunications, China, in 2007 and 2010, respectively. From 2011 to 2016, she was a lecturer at Qijing Normal University, China. Since 2017, she has been an associate professor at Qijing Normal University. Her research interests are in remote sensing, wireless navigation, and artificial intelligence. (tianah@mail.qjnu.edu.cn)