A New Semi-supervised Classification Method of Hyperspectral Image based on Combining Renyi Entropy and Multinomial Logistic Regression Algorithm

Chunyang Wang¹, Shuangting Wang¹, Zengzhang Guo¹*, Liping Wang² and Chao Ma¹

¹School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, 454000, China;
²School of Computer Science, Engineering and Mathematics, Flinders University, South Australia 5042, Australia
hpu.wcy@gmail.com, gzc@hpu.edu.cn

Abstract

The study on classification methods of hyperspectral image is a focal growing area in remote sensing applications because the wide spectral range, providing a very high spectral resolution, allows the detection and classification surfaces and chemical elements of the observed image. Semi-supervised learning method which takes a large number of unlabeled samples and minority labeled samples, improving classification and predicting the accuracy effectively have been a new research direction. In this paper we proposed a new semi-supervised classification method of hyperspectral image based on combining Renyi entropy and multinomial logistic regression algorithm. The multinomial logistic regression was performed to describe a direct relationship between the selected sample and their category. A lot of unlabeled samples are constantly added to the sample data using Renyi entropy algorithm. The test analysis of image classification in test area showed the advantages of classification method based on combining Renyi entropy and multinomial logistic regression algorithm for hyperspectral remote sensing image.

Keywords: hyperspectral image, image classification, multinomial logistic regression, Renyi entropy, semi-supervised learning

1. Introduction

In recent years, analysis of hyperspectral images have been attracted more and more attentions in many fields such as military and civilian. The advantage of high spectral resolution is that it can obtain many continuous band spectral images when observing. So it achieves a fine description of ground targets and reaching the purpose of identifying features [1-2].

Currently, supervised classification and unsupervised classification are the two traditional classification methods for remote sensing image. The supervised classification is based on class probability density function for samples in spatial feature. Generally, it takes higher classification accuracy and needs a lot of correct training samples. Unsupervised classification is a clustering method and can detect unknown classes on images. The advantage of unsupervised classification is simplify and efficiency. However, it can’t guarantee the real relationship between the clustering features classes and surface features classes [3-4]. Semi-supervised learning method takes a large number of unlabeled samples...
and minority labeled samples, improving classification and predicting the accuracy effectively, so that it overcomes the shortcomings between supervised classification and unsupervised classification effectively [5-6].

The study of using a small amount of labeled samples and selecting the useful information to help estimate data classification becomes a new hotpot based on semi-supervised learning research for hyperspectral image classification [7]. Li, et al., (2010) proposed the method of choosing unlabeled samples; minimum prediction probability was employed to approximate maximum entropy to add unlabeled samples to training data sets [8]. Yao, et al., (2012) conducted the following research, that the constraint of spatial neighborhood was imposed into the kernel function of Gaussian process, so the spatial correlations of labeled and unlabeled samples can be embedded in the kernel function to supple the training data of hyperspectral image [9]. Wang, et al., (2012) reported the usage of multiple adjacent point information in the neighborhood to train unlabeled sample data through k-Nearest-Neighbor Simplex algorithm [10]. Huang, et al., (2014) adopted the Sparse Discriminant embedding algorithm to train the unlabeled samples and got the intrinsic attribute of high dimensional hyperspectral image data and the manifold structure of low dimensional data [11]. There are some drawbacks, although these methods achieve the purpose of the classification for hyperspectral image. One kind of ordinary algorithm is to choose the unlabeled samples to add them to the training set through random selection strategy. The misjudgment probability of this strategy is relatively big. It can lead to error accumulation in the iterative process. Another kind of strategy is to select the unlabeled samples that have the smallest judging probability for each point in the pre-iterative process. In the other words, it was equivalent to the selection of maximum entropy unlabeled samples approximately and added them to the training set, then conducted the next iteration. The accuracy of this strategy is higher than the accuracy of random selection strategy, but it still cannot improve classification accuracy significantly especially the type classified is many.

Renyi entropy was presented as the form of information entropy generalization in 1961 [12], it was used in many areas in initially, such as adaptive filtering, signal communication system balancing and fault detection machinery [13-14]. Renyi entropy was the standard for information content measurement and could detect the richness of global or local information content effectively. In recent years, the application to processing and analyzing remote sensing image was becoming a hot topic. Wang, et al., (2012) used Renyi entropy to reduce dimension and extracted principal components using combining with the SVM algorithm. The hyperspectral image data was classified and achieved better results [15]. Hu, et al., (2013) combined normalization and Renyi entropy to match the control point on images automatically, and got a high accuracy [16].

In the present study, a semi-supervised classification method was proposed for hyperspectral remote sensing image classification based on combining Renyi entropy and multinomial logistic regression algorithm. The core idea of this algorithm is that, firstly, selected small amount of sample data are performed by multinomial logistic regression algorithm. The fitted regression coefficient can describe the direct relationship between selected sample pixel and its category effectively. Then, hyperspectral image is classified by using the fitted regression coefficient. Secondly, the entropy of the experimental area is calculated through Renyi entropy calculation method, and the some unlabeled samples of maximum Renyi entropy are selected from the calculation data to add to the sample data. The classification of multinomial logistic regression is not iterated repeatedly for many times until the classification accuracy tends to be stable. Finally, the paper has classification experiment by the real AVIRIS hyperspectral remote sensing image data. It shows that it can improve the accuracy of classification and obtain better classification results by using a new method and it
provided a new thought for the study of the classification of hyperspectral remote sensing image.

2. Materials and Methods

2.1. Problem Description

For hyperspectral classification problems, assuming the hyperspectral remote sensing image is \( X = (x_1, \ldots, x_n) \in \mathbb{R}^{d \times n} \), \( d \) represents the number of bands, each pixel represents a vector, there are \( d \) observations, \( C = \{1, \ldots, C\} \), \( C \) represents category set, \( T = \{x_1, \ldots, x_n\} \) shows that training set contains sample labels, \( \tau \subseteq x \), \( U \) represents unlabeled sample set, \( X = \tau \cup U \), \( Y = (y_1, \ldots, y_n) \in C \) represents classification categories of sample training datasets.

2.2. The Principle of Multinomial Logistic Regression Algorithm

The multinomial logistic regression algorithm can predict the fitted coefficient quickly and accurately [17-18], the vector of parameter coefficient \( \beta = (\beta_1, \ldots, \beta_n) \) is got by using multinomial logistic regression algorithm for labeled training sample set. Equation 1 represents the probability formula that probability of the event occurring that multinomial logistic model is expressed as outcome variables.

\[
P(y_i = j | x_i) = \frac{\exp(x_i \cdot h(x_i))}{1 + \sum_{j=1}^{c} \exp(x_i \cdot h(x_i))}
\]  
(1)

When \( j \) is equal to \( 1, \ldots, K \), the \( h(x_i) \) can be represented as the Equation 2

\[
h(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon
\]  
(2)

The coefficient of \( \beta \) is estimated through the estimation criteria of Bayesian maximum a posteriori and the log-likelihood algorithm [13-14], Equation 3 gives the solving method.

\[
w = L(\beta) = \sum_{i=1}^{n} \left[ \log(\exp(h(x_i))) \right]
\]  
(3)

The Equation 3 calculates fitted coefficient of multinomial logistic classification, the estimating type \( CLA(x_{k+1}) \) brought by a new pixel \( x_{k+1} \) is calculated through the Equation 4.

\[
CLA(x_{k+1}) = \arg \max_w \sum_{i=1}^{c} w
\]  
(4)

Each predicted category \( (CLA_1, \ldots, CLA_n) \) for hyperspectral remote sensing image is calculated through Equation 4, it represents a classification process is completed.

The important issue studied is how to add new samples to the training sample in the process of semi-supervised classification. This process is conducted automatically by predicting through the current classifier. The samples get through predicting by multinomial logistic regression classifier, which have the biggest probability was added. Although sample category is exactly, the amount of information contained is least. The positive effect on the next classification process is the smallest and it can’t improve classification accuracy if these samples are added to the training set. On the contrary, it will increase the computational burden of the classification algorithm in the process of the selected training samples. So the research for selecting unlabeled sample added to the training sample is very important.
2.3. Selected Unlabeled Samples using Renyi Entropy Algorithm

Renyi proposed Renyi entropy in 1961[12], the definition form is shown in Equation 5.

\[
R_{\alpha}(X) = \frac{1}{1-\alpha} \ln \left( \sum_{i=1}^{n} p(x_i)^{\alpha} \right), \quad \alpha \geq 0, \quad \alpha \neq 1
\]  

(5)

Renyi entropy can reflect the uniformity of the attribute value distribution, and it is a scale for measuring the degree of uncertainty for information and the amount of information. This paper uses renyi entropy with two times to describe the amount of information. It is shown in Equation 6.

\[
R_{2}(X) = -\ln \left( \sum_{i=1}^{n} p(x_i)^{2} \right), \quad \alpha = 2
\]  

(6)

where \( p(x_i) \) takes the predicted probability value of each pixel because the object of study is hyperspectral image data. The Equation 6 is normalized to prevent infinite value appeared in the process of calculation and obtained the Equation 7.

\[
R_{2}(X) = -p(x_i) \ln \left( \sum_{i=1}^{n} p(x_i)^{2} \right)
\]  

(7)

Hyperspectral image data set contains a large amount of implicit information, through the comprehensive analysis of the statistical information in the data set. Some regularity knowledge is used to obtain data connection. The classification result of sample that has the bigger entropy is uncertain for the current classifier and is the most informative sample. The new training set which is retrained with the unlabeled samples of maximum Renyi entropy is used for the new classification process. The overall accuracy of hyperspectral image classification is greatly improved. \( X = (x_1, ..., x_n) \in \mathbb{R}^{d \times n} \) represents the hyperspectral image data, the maximum Renyi entropy value is calculated by the Equation 8.

\[
S(U) = \max \{ R_{2}(X) \}
\]  

(8)

\( S(U) \) represents the final choice of unlabeled set in the above formula, \( R_2(X) \) represents Renyi entropy calculated for each pixel on image in the process of the last classification, and then values of Renyi entropy are sorted. The \( k \) samples extracted from them are added to the existing training sample set. The \( k \) is the number of label in supplementary training set.

2.4. The Process of Algorithm

The process of the algorithm is shown, as follows:

Input: the original hyperspectral image \( X = (x_1, ..., x_n) \in \mathbb{R}^{d \times n} \), the selected training sample set \( T = \{x_1, ..., x_i\} \), \( T \subseteq X \), the corresponding classification categories of training set \( Y = (y_1, ..., y_j) \in C \), unlabeled sample set \( U \), \( X = T \cup U \). \( U^{\text{new}} \) Represents the number of selected new unlabeled samples every time. The \( m \) represents the number of semi-supervised training, and \( x \) \( r_{U^{new}} \) represents the remained unlabeled tag on image.

Output: the classification result of the hyperspectral image data.

The process is described as below:

Step1: Initialize setting and loading original image \( X = (x_1, ..., x_n) \in \mathbb{R}^{d \times n} \), and structure the matrix of the image.
Step2: Load the training set \( T = \{ x_1, \ldots, x_k \} \) and the categories corresponding to training set \( y = (y_1, \ldots, y_k) \).

Step3: Compute regression coefficient \( \beta \) on the basis of the Eq. (1), (2) and (3).

Step4: Use \( \beta \) to structure the model using the proposed algorithm in this paper of predicting classification.

Step5: Select unclassified pixels \( x_{r \in U_{\nu}} \) in image \( X \), and take them in the Eq. (4), and predict the classification result \( (CLA_1, \ldots, CLA_n) \).

Step6: Based on Eq. (6), calculating the Renyi entropy value of each pixel in \( x_{r \in U_{\nu}} \).

Step7: Based on Eq. (7), the new \( k \) pixels of unlabeled sample set \( U_{\nu} \) of maximum Renyi entropy were extracted and used to update training set \( T \). \( T_{\nu} = x_{r \in U_{\nu}} + x_{r \in U_{\nu}}, p=(1 \ldots m) \)

Step8: Return to step (3), and iterate \( m \) times of the semi-supervised learning process.

Step9: Output the classification result of the hyperspectral image, and the process of algorithm is over.

3. The Study Area and Validation Images

The typical test images and the parameters of which are presented in Table 1. The T-1 data is the NASA AVIRIS instrument acquired data over the Kennedy Space Center (KSC), Florida, on March 23, 1996. The data in 224 bands of 10nm width with center wavelengths from 400–2500nm have a spatial resolution of 18 m. In the present study, 120 bands were selected to analysis from the study area data which is removing the effects of water vapor absorption and low SNR bands. The selection of training data is provided by Landsat Thematic Mapper images. According to the interpretation of images, the land cover within the experimental area is divided into 13 major categories [19], meanwhile, choosing the different number of training sample points to each category. The image parameters of experimental zone are listed in Table 1, then in the Figure 1 (a) is the true color image of the experimental area image, the following shown in Figure 1 (b) is the training samples of each category feature, in the Figure 1 (c) is the spectral curve of 13 kinds of major categories feature and in the last one Figure 1 (d) is the thematic maps definition of 13 classes.

<table>
<thead>
<tr>
<th>Image</th>
<th>Platform</th>
<th>Size (pixel)</th>
<th>Spatial resolution (m)</th>
<th>Band composition</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-1</td>
<td>AVIRIS</td>
<td>614 x 512</td>
<td>18</td>
<td>224 bands, covering 400-2500nm</td>
<td>KSC, Florida, USA</td>
</tr>
</tbody>
</table>
4. Results

4.1. Classification Accuracy Comparison between Different Label Selection Algorithms

Different methods are selected for choosing label to compare classification results, and it can illustrate the effectiveness of the algorithm. The methods are random selection method, minimum probability method and maximum Renyi entropy method. Since the prediction coefficient in the initial process multinomial logistic regression has certain randomness. In order to ensure the accurate result, the test for each method is conducted 10 times, and the average results are taken. The accuracy rate curve of average classification is shown in Figure 2.

![Figure 2. The Accuracy Rate Curve of Average Classification](image)

Figure 2 shows that the differences between results of the classification are not obvious among label selection methods in the initial stages of the iterative process. The reason for the above phenomenon is that the number of classifier training sample is less and the classifier
trained is not accurate in the initial stages. But as the iteration is constantly running, the role of measuring information of the maximum Renyi entropy is appearing gradually. The adopted classification method is better than the other two methods from the overall accuracy. When the number of iteration is same (fixed x-axis), the highest accuracy of classification is obtained using the Renyi entropy method. On the other hand, with the number of training sample increasing, the stable result which the classification accuracy rate tends to be smaller, and reaches a critical state eventually can also be proved from Figure 2.

4.2. Classification Accuracy Comparison for Different Classifier Algorithms

To illustrate the effectiveness of the algorithm, the several representative machines learning classification algorithms including supervised algorithm and unsupervised algorithm are selected to compare with the new algorithm proposed. The classification results of variety classification algorithms are shown in Figure 3. The classification results of unsupervised classifier of the k-means algorithm are shown in Figure 3(a), The classification results of supervised classifier of the minimum distance algorithm are shown in Figure 3(b), the classification results of supervised classifier artificial neural network algorithm are shown in Figure 3(c), the classification results of supervised classifier of support vector machine algorithm are shown in Figure 3(d), and the final Figure 3(e) is the classification results of the semi-supervised classifier of the proposed algorithm in this paper. Table 2 is a comparison of various classification algorithm running time, overall accuracy and Kappa coefficient evaluation indexes.

![Figure 3. The Classification Results of Study Area using Various Classifier Algorithm](image)


The results of comparison are shown in Figure 3 and Table 2. Based on the classification results of various classification algorithms, it can be concluded that the overall classification accuracy of unsupervised classifier of the k-means algorithm is lowest although the running time is shortest. The serious leak misclassification is shown Figure 3 and the only six classes were distinguished from unsupervised classifier.
### Table 2. Overall Accuracy and Kappa Coefficient based on Comparison of the Vary Classification Methods

<table>
<thead>
<tr>
<th>Classification methods</th>
<th>Iteration Time(s)</th>
<th>Overall Accuracy(%)</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means classifier</td>
<td>30</td>
<td>53.07%</td>
<td>0.51</td>
</tr>
<tr>
<td>minimum distance classifier</td>
<td>205</td>
<td>64.46%</td>
<td>0.61</td>
</tr>
<tr>
<td>neural network classifier</td>
<td>491</td>
<td>67.01%</td>
<td>0.64</td>
</tr>
<tr>
<td>support vector machine classifier</td>
<td>165</td>
<td>84.41%</td>
<td>0.83</td>
</tr>
<tr>
<td>combining Renyi entropy and multinomial logistic regression classifier</td>
<td>375</td>
<td>95.92%</td>
<td>0.94</td>
</tr>
</tbody>
</table>

It can be concluded that the overall classification accuracy of supervised classifier of minimum distance algorithm is better than the unsupervised k-means classifier algorithm. But it is low due to the moderate running time and more serious leak misclassification. For instance, it can be clearly seen from the Figure 3(b) that the CP/Oak feature almost disappeared in the classification diagram, even if the place where marked the training sample is also hard to find.

The overall classification accuracy of artificial neural network classifier algorithm is also relatively low, because of longer running time, inefficient and more serious misclassification error. Figure 3(c) shows that the surface Water feature of the mixed pixel that nearby the waterfront cannot be distinguished by the algorithm, and made the mixture of surface features such as CP/Oak feature, Slash Pine feature, Cattail Marsh feature looks like the Water feature.

The quality of support vector machine classifier algorithm is relatively high. The advantages of this algorithm include the running speed, the relatively high overall classification accuracy and the best effect of classification for a large contiguous area. However, this algorithm reveal the low capacity for analyze the small surface features appear the error of leakage. For instance, the SVM algorithm will eliminate the small building in the water body.

This paper develop a classier base based on combining Renyi entropy and multinomial logistic regression algorithm which has several advantages like the relatively high overall classification accuracy, less leakage points and misclassification error. It is better than the other types of classification algorithms especially for many types of dense regions, but the drawback is particularly serious for long-running time and need further more computing power and hardware.

### 5. Conclusions and Future Work

In this study, a new semi-supervised method of classification for hyperspectral remote sensing images was proposed, especially focuses on semi-supervised machine learning application using the combining Renyi entropy and multinomial logistic regression algorithm. Firstly, according to the features of hyperspectral data, the multinomial logistic regression algorithm was employed to predict the primary classification results based on the selected few labels in training sets. Secondly, the Renyi entropy algorithm was used to select the pixel of
maximum information in hyperspectral image to add to the training samples dataset. Finally, the proposed method iterates the selection and learning processes to output the final classification result. The classification test of study area image was conducted using the new algorithm. The experimental results showed that the algorithm has a high precision overall classification, less misclassification errors and leakage points, and more advantages at intensive more multiclass region classification.

However, the new algorithm still has inadequacies, such as long-running time and need further more computing power and hardware. Future research priority will focus on optimizing algorithm, saving the running time, and promoting working efficiency.

Acknowledgements

This research is jointly supported by the National Natural Science Foundation of China & Shenhua Group Corporation Limited (U1261206). The authors would like to thank Dr. Jun Li at Technical University of Lisbon (Portugal) for kindly sharing her demo code and Dr. Jianghui Dong at Flinders University (Australia) for greatly helping to improve the quality of this paper.

References


Authors

Chunyang Wang, he received his B. Eng. degree from Changchun University of Science and Technology, Changchun, China (2002) and M. Sci. Degree from Northeast Normal University, Changchun, China (2006). He is currently pursuing the Ph.D. degree in School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include image processing, classification method of hyperspectral image, machine learning and pattern recognition of remote sensing application.

Shuangting Wang, he is a full Professor in School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include photogrammetry, remote sensing image processing, remote sensing information extraction.

Zengzhang Guo, he is a full Professor in School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. He received his Ph.D. degree from China University of Mining and Technology, Beijing, China (2000). His research interests include mining subsidence monitoring and remote sensing application, protection of mining-induced ground damage, spatial data acquisition and processing.

Liping Wang, she received her M. Eng. degree in mechanical engineering from Tianjin University, Tianjin, China (2005). She is an Associate Professor in Shanghai Jianqiao University, Shanghai, China. Her research interests include mechanical design, mechanical image processing, reverse engineering, bio-mechanics and biomaterials.
Chao Ma, he is an Associate Professor in School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China. He received his Ph.D. degree from State Key Laboratory of Earthquake Dynamics, Institute of Geology, China Earthquake Administration, Beijing, China (2005). His research interests include mining subsidence Interferometry of Synthetic Aperture Radar (InSAR), vegetation ecological changes and microwave scattering mechanism of mining area of remote sensing application.