

# Affinity Propagation for Hyperspectral Band Selection

Zhiwei Ren <sup>1,\*</sup>, Lingda Wu <sup>2</sup>

<sup>1</sup>Department of Graduate Management, Space Engineering University, Beijing 101416, China

<sup>2</sup>Science and Technology on Complex Electronic System Simulation Laboratory Space Engineering University, Beijing 101416, China

\*Corresponding author e-mail: juimer@foxmail.com

**Abstract.** Because hyperspectral images have the characteristics of high correlation between bands and strong information redundancy, the reduction in dimension of hyperspectral images is an important step in the pre-processing of hyperspectral images. Band selection can preserve the physical meaning of the original data while reducing dimension and has application in many aspects. Affinity Propagation Clustering (AP) is a clustering method proposed by Fray et al. in 2007. AP clusters based on the correlation between data points and treats all data points as potential cluster centers. This paper proposes a band selection method based on AP clustering, which introduces wavelet transform into the calculation of similarity and preference value in clustering algorithm. The dimensionality reduction results are input into the minimum distance classifier for classification, and the classification accuracy was calculated. The dataset is validated by the Indiana Pines dataset. The experimental results verify the effectiveness of the proposed method.

## 1. Introduction

The hyperspectral image can not only describe the spatial shape and distribution of the ground features, but also has the characteristics of high spectral resolution, continuous spectrum, and one-dimensional map. However, the spectral resolution of hyperspectral images is too high, resulting in problems such as large data volume, strong correlation between bands, and serious information redundancy. Therefore, the difficulty of hyperspectral image processing is much higher than that of color images and multi-spectral images. And the "Hughes" phenomenon is obvious. In order to avoid possible dimensional disasters and improve processing efficiency, it is often used to reduce dimensions before hyperspectral image classification analysis.

At present, there are mainly two methods for dimension reduction of hyperspectral images: spectral feature extraction and spectral feature selection. Spectral feature selection, also known as band selection, selects a subset of the spectral feature space for a particular object. Feature extraction refers to the process in which the original spectral feature space or its subspace achieves the purpose of dimensional reduction, feature enhancement and the like through some mathematical transformation.

The purpose of band selection is to select a subset of bands or bands that are large in information, weak in correlation, and representative. Common band selection methods can be divided into unsupervised, supervised, and semi-supervised band selections based on whether labeled sample information is used. The supervised band selection uses the separability of the marker samples to select



the band subsets. The unsupervised band selection gives a certain index by examining the size of the information contained in the band and the correlation between the bands. According to the given index, all bands are sorted in descending order, and the first few bands satisfying the requirements are selected. The methods of giving indicators include the best exponential factor method, adaptive band selection method, entropy and joint entropy method, and automatic subspace partitioning method. Semi-supervised band selection can make full use of limited labeled sample data and a large amount of unlabeled sample data, which can not only obtain higher classification accuracy, but also meet the requirements of real hyperspectral data. It has attracted more and more attention of researchers. At present, there are many semi-supervised band selection methods based on self-training and maps.

Band selection based on clustering is also a feasible way to implement band selection. The clustering method divides the samples in the data set into a number of usually disjoint subsets. The clustering center of each subset is usually the most representative sample in the subset. Therefore, it is feasible to use the clustering method for band selection. However, the traditional clustering method is sensitive to the initial clustering center. It needs to be considered that the clustering number is determined, and the randomness of the initial center selection may cause instability of the clustering result. In addition, the result of traditional clustering is not a real band, so there is a big error between the final band selection result and clustering. In 2007, Fray et al. proposed a nearest neighbor propagation clustering algorithm (AP) to solve the problem that the clustering result is sensitive to the initial clustering center and does not need to achieve the specified number of clusters; it is based on data points. The correlations are clustered and all data points are regarded as potential clustering centers, so they are not affected by the initial center. Compared with traditional clustering methods, AP algorithm has good clustering effect when dealing with multi-class and large-scale data, and the algorithm is relatively stable. At present, AP algorithm has been successfully applied to image segmentation, target recognition and other fields.

Based on AP clustering, this paper uses wavelet decomposition to process hyperspectral images. The resulting high frequency component contains details of the image and noise information, and the resulting low frequency component contains image smoothing information. The high-frequency components are used to calculate the correlation between each band and the signal-to-noise ratio to complete clustering. This paper proposes an algorithm that can select the band with high signal-to-noise ratio and weak correlation. The obtained clustering results are input into the minimum distance classifier for classification, and the effectiveness of the algorithm is verified.

## 2. AP clustering improved by Wavelet Transform

### 2.1. AP Clustering Algorithm

Assume that dataset  $D$  has  $n$  samples  $\{x_1, x_2, \dots, x_n\}$ . The AP algorithm first computes the similarity  $S$  between each two samples and calculates the degree of attraction  $R$  (responsibility) and the degree of attribution  $A$  (availability) according to the similarity, combining the degree of attraction and attribution. Degree of information in two areas to find the best set of class representative points, and ultimately make the sum of the similarity of all data points to their nearest class representative point.

The similarity matrix  $S$  is the basis of the work, each element is a negative value, and the corresponding index can be selected as a measure according to different tasks. When  $i = j$ ,  $s(i, j)$  represents bias parameter  $p(i)$ .  $r(i, j)$  denotes the degree of attraction of the data point  $j$  to the point  $i$ , and describes the degree to which the point  $j$  is suitable to represent the class of the point  $i$ .  $a(i, j)$  indicates that the data point  $i$  to point  $j$  is the degree of attribution, indicating the degree to which data  $i$  selects point  $j$  as the class representation. The larger  $r(i, j)$  and  $a(i, j)$ , the greater the probability that point  $j$  is the final cluster center. The AP algorithm continuously collects and transmits evidence (also known as message passing) through an iterative loop to generate  $m$  high-quality class representations and corresponding clusters. The energy function of the clustering is also minimized. The

points are assigned to the class that the nearest class represents belongs to, then the  $m$  clusters found are the clustering results.

The formulas for the attraction matrix and the attribution matrix are as follows.

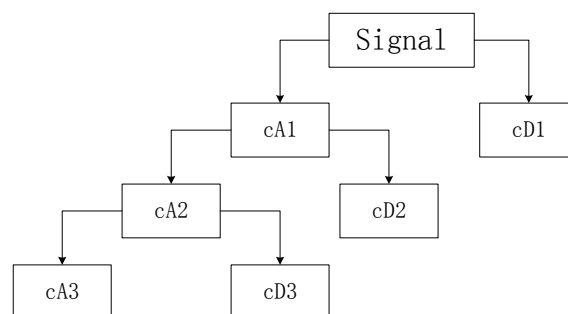
$$r(i, j) = \begin{cases} s(i, j) - \max_{j \neq j'} \{a(i, j') + s(i, j')\}, & i \neq j \\ p(j) - \max_{j \neq j'} \{a(j, j') + s(j, j')\}, & i = j \end{cases} \quad (1)$$

$$a(i, j) = \begin{cases} \min\{0, r(j, j) + \sum_{i' \notin (i, j)} \max(0, r(i', j))\}, & i \neq j \\ \sum_{i \neq j} \max\{0, r(i', j)\}, & i = j \end{cases} \quad (2)$$

## 2.2. AP clustering improved by Wavelet Transform

Band selection is a pre-processing result in hyperspectral image processing. Band selection results influence the classification and accuracy of target recognition. In addition to the amount of information contained in the band and the inter-band correlation, noise size is also an important factor affecting accuracy. At present, there are many methods used to calculate the correlation between spectra, including distance measures, encoding measures, and so on. In addition, through the corresponding transformation of the data, correlation calculations such as wavelet transforms are performed. Calculating the SNR of an image requires separating the noise from the signal. The wavelet transform can divide the image into high-frequency components containing details and noise information and low-frequency components including smooth information such as the feature background. Therefore, in this paper, wavelet transform is used to calculate the correlation between bands and the signal-to-noise ratio of the band, and AP clustering algorithm is improved to achieve band selection.

Multi-scale wavelet transform is used as a one-dimensional signal of the spectral curve to obtain low-frequency components and high-frequency component vectors at different scales. The similarity of the original spectral vector is analyzed by measuring the similarity of the low-frequency and high-frequency vectors. Because the low-frequency component mainly contains image smoothing information, it can reflect the approximate characteristics of the spectral curve; the high-frequency component reflects the detailed features of the spectral curve and noise information. Therefore, the low-frequency components of each curve have higher similarity, and the similarity of high-frequency components is low. Using high-frequency components for similarity analysis will achieve more objective results. In this paper, three-layer Sym4 wavelet decomposition is performed on the spectral curve, and the similarity measure is performed on high frequency components.



**Figure 1.** Three-level wavelet decomposition

For the high-frequency components of the wavelet transform, this paper uses the SAM to calculate the correlation between the two bands and constructs the similarity matrix  $s$  for AP clustering. SAM's formula is as follows.

$$s(x, y) = \arccos\left\{\frac{\sum_{i=1}^L s_x^i s_y^i}{[\sum_{i=1}^L (s_x^i)^2]^{1/2} \times [\sum_{i=1}^L (s_y^i)^2]^{1/2}}\right\} \quad (3)$$

Among them,  $s_x$  and  $s_y$  indicate two bands for calculating similarity;  $L$  indicates the number of pixels in the band.  $s^i$  represents the pixel value of the  $i$ th pixel.

After the wavelet transform, most of the energy of the image is concentrated on the low-frequency component. The high-frequency component has a small amplitude and a low energy. Therefore, when the noise is large, the coefficients of the highest frequency subband can all be considered to be noise, thereby estimating the standard deviation of the noise. Donoho and Johnstone proposed the estimation formula of the noise standard deviation in the wavelet domain, the formula is as follows.

$$\sigma_n = MAD / 0.6745 \quad (4)$$

Among them,  $MAD$  is the median of the wavelet coefficient amplitude of the high-frequency component.

Therefore, the signal-to-noise ratio calculation formula for the  $a$ th band of the hyperspectral image is,

$$SNR_i = 10 \times \log_{10}(\sigma_{si}^2 / \sigma_{ni}^2) \quad (5)$$

Among them,  $\sigma_{si}$  and  $\sigma_{ni}$  respectively represent the signal standard deviation and noise standard deviation of the band.

### 3. Band Selection Based on AP Clustering

In connection with the previous section, the process of wavelet selection (WDAP) based on wavelet transform and improved AP algorithm is as follows.

Input: normalized hyperspectral image data ( $N$  bands)

Output: One-dimensional vector  $idx_{1 \times k}$ , where  $k$  represents the number of bands selected by the band, and  $idx_{1 \times j}$  represents the label of the band where the cluster center is located,  $j = 1, 2 \dots k$ .

Step 1 performs wavelet transform on each band of hyperspectral image data in sequence;

Step 2 proposes that the high frequency component is calculated according to equation (3), and the preference value  $p$  is calculated according to equations (4) and (5);

Step 3 For AP clustering, see section 1.1.

The algorithm does not rely on the real tag of the feature and belongs to unsupervised band selection. At the same time, the clustering algorithm does not depend on the setting of the initial category center and has a large practical value.

### 4. Experimental verification

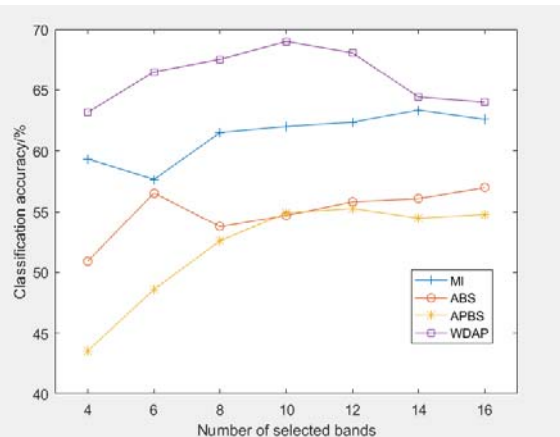
The method proposed in this paper (WDAP) and the maximum information method (MI) proposed in the literature [10], the automatic subspace division method (ABS) proposed in the literature [6], and the band selection method (APBS) based on unimproved AP clustering Comparative Test. The output

results of each band selection method are input into the minimum distance classifier for classification processing, and the classification accuracy and running time are compared and analyzed.

Experiments were performed on the Intel(R) Core(TM) i7-6500U CPU and the 8192MB RAM hardware platform using MATLAB software. The Indian Pines data set, which included 16 feature categories, collected by an on-board imaging spectrometer AVIRIS in an agroforestry mixing room in northwestern Indiana, USA, was used for the experiment. The spatial resolution of the image is 25m, and the image size is pixels. The original data has 224 spectral bands in the wavelength range. Finally, 200 bands with high signal-to-noise ratio and good quality are reserved. The Indian Pines data set grayscale image is shown in Figure 2.



**Figure 2.** Indian Pines Dataset



**Figure 3.** Comparison of classification accuracy

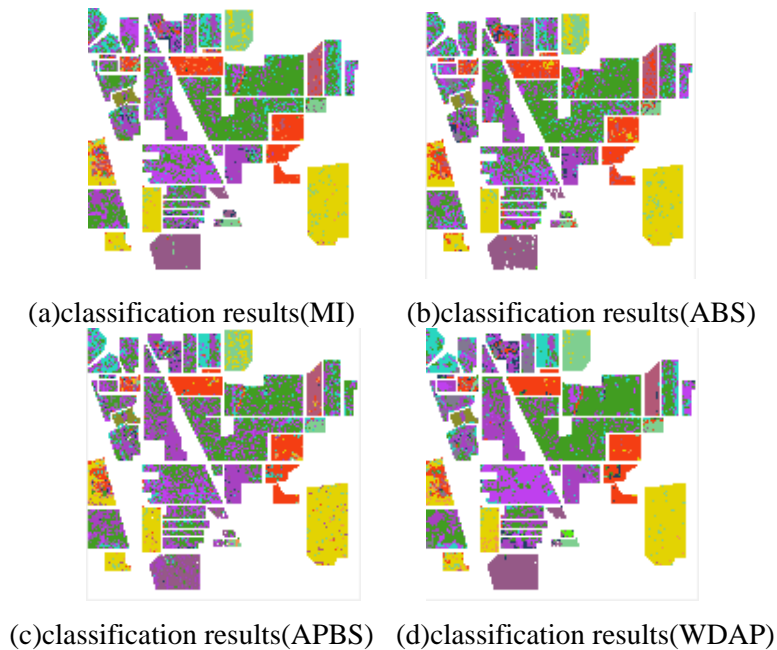
Figure 3 is a graph showing the overall accuracy (OA) of the above four band selection algorithms. It can be seen from the curves that the overall classification accuracy of the proposed WDAP is significantly higher than the other three band selection methods, especially when the number of bands is less than 10. This shows that WDAP can fully exploit the effective information of hyperspectral data. The accuracy of the ABS and APBS methods is low. When the number of bands is higher than 10, the accuracy of the WDAP method slightly decreases, but it is still higher than the other methods. At this time, the classification accuracy of MI, ABS and APBS changes slowly. This shows that when the number of bands increases to a certain extent, the increased bands cannot provide more effective information that is conducive to classification, and invalid or even harmful information increases. It is reflected that the information redundancy between the bands of the hyperspectral data image is serious. Once again proved the necessity and effectiveness of dimensionality reduction.

Table 1 compares the running time when the number of selected bands is 10. As can be seen from the table, the MI method running time is significantly higher than the other three methods. The ABS method has a lower running time because the method only considers the correlation between adjacent bands, while the remaining methods consider the correlation between the two bands. The WDAP runs longer than the APBS method because WDAP adds time for wavelet calculations and spectral angle mapping. However, the WDAP method takes much less time than the MI method. Therefore, the feasibility of the proposed method in practical applications is strong. In the next step of research, we will focus on the study of a more efficient method of calculating the degree of preference instead of the spectral angle mapping method in the text to shorten the running time of the algorithm.

**Table 1.** Comparison of running time of different methods (10 bands)

Algorithm name	MI	ABS	APBS	WDAP
Algorithm runtime /s	42.61	0.52	2.74	8.38

Figure 4 (a ~ d) is a schematic diagram of the classification of four methods of ground objects at 10 bands. As can be seen from the figure, the WDAP method has obvious advantages for the classification of areas with large areas and less details. The performance in the more detailed areas has yet to be strengthened. When the wavelet transform is introduced in the algorithm and the signal-to-noise ratio is calculated, partial detail errors are calculated as noise. The next study will improve this issue.



**Figure 4.** Classification results of the four band selection methods

## 5. Conclusion

This paper proposes a method of band selection for hyperspectral images based on improved AP clustering based on wavelet transform. The spectral similarity is calculated by using spectral angle mapping, and the degree of preference is calculated by using signal to noise ratio. The Indiana Pines data set was dimensionally reduced by designing experiments, and the dimensionality reduction results were input into the minimum distance classifier for feature classification. Experiments show that the accuracy of band classification for hyperspectral image band selection based on improved AP clustering based on wavelet transform is higher than that of MI, ABS, APBS and other methods. Compared with the above methods, the hyperspectral image band is improved based on wavelet transform and improved AP clustering. Select information that can use data sets more efficiently to improve the accuracy of the classification. And the increase in calculation costs is within a reasonable range. In the future research, other advanced similarity calculation methods will be used to improve the efficiency and accuracy of the algorithm.

## References

- [1] G, Hughes. On the mean accuracy of statistical pattern recognizers[J]. Information Theory IEEE Transactions on, 1968, 14(1): 55-63.
- [2] Zhang Liangpei, Zhang Lifu. Hyperspectral Remote Sensing [M]. Beijing: Surveying and Mapping Press, 2011.
- [3] SU Hong-jun, DU Pei-jun, SHENG Ye-hua. Study on band selection algorithm for hyperspectral imagery[J]. Research on Applications of Computers, 2008, 25(4): 1093-1096.
- [4] Lu Zhiqiang. A Dissertatio Submitted to Shanghai University for the Degree of Master in Engineering [D]. Shanghai: Shanghai University, 2013.
- [5] Du P, Du Q. Semi-supervised dimensionality reduction using orthogonal projection divergence-



- based clustering for hyperspectral imagery[J]. Optical Engineering, 2012, 51(11):1715.
- [6] Liu Chunhong, Zhao Chunhui, Zhang Lingyan. A New Method of Hyperspectral Remote Sensing Image Dimensional Reduction [J]. Journal of Image and Graphics, 2005, 10(2): 218-222.
- [7] Xia Chunmeng, Ni Zhiwei, Ni Liping, Zhang Lin. Neighbour Propagation Algorithm Based on Density Adjustment and Manifold Distance[J]. Computer Science, 2017, 44(10): 187-192.
- [8] Donoho D L, Johnstone I M. Ideal Spatial Adaptation Via Wavelet Shrinkage [J]. Biometrika, 1994, 81(3): 425-455.
- [9] Wang Jing, Wang Lijun, Cui Jiantao, Li Xiaorun. Band Selection based on Signal-to-noise Ratio Estimation and Hyperspectral Anomaly Detection[J]. Remote Sensing Technology and Application, 2015, 30(2): 292-297.
- [10] Liu Xuesong, Ge Liang, Wang Bin, Zhang Liming. An unsupervised band selection algorithm for hyperspectral imagery based on maximal information [J]. Journal of Infrared and Millimeter Waves, 2012, 31(2): 166-170.