

HYPERSPECTRAL BAND SELECTION BASED ON IMPROVED AFFINITY PROPAGATION

Qingyu Zhu¹, Yulei Wang^{1,2}, Fengchao Wang¹, Meiping Song¹, Chein-I Chang^{1,3}

¹ Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China;

² State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xian 710071, China

³ Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

ABSTRACT

Dimensionality reduction is a common method to reduce the computational complexity of hyperspectral images and improve the classification performance. Band selection is one of the most commonly used methods for dimensionality reduction. Affinity propagation (AP) is a clustering algorithm that has better performance than traditional clustering methods. This paper proposes an improved AP algorithm (IAP), which divides each intrinsic cluster into several subsets, and combines the information entropy to change the initial availability matrix to obtain a suitable number of clustering results with arbitrary shapes. The experimental results on the public hyperspectral data set show that the band combination selected by IAP has a better classification accuracy compared with all bands data set and band subset by traditional AP algorithm.

Index Terms — Hyperspectral images, affinity propagation, improved affinity propagation, classification

1. INTRODUCTION

Hyperspectral images contain higher spatial resolution and spectral resolution, providing rich information for the classification and identification of ground features [1]. However, when the training samples are limited, the classification accuracy will decrease as the band increases, which is the "Hughes" phenomenon. In addition, the correlation between adjacent bands is relatively high, causing information redundancy and increasing the amount of calculation. Therefore, dimensionality reduction is necessary.

At present, the dimensionality reduction of hyperspectral remote sensing images is mainly divided into two methods: one is feature extraction based on transformation. The original data is spatially transformed through mathematical mapping and transformed to a

low-dimensional space. The new feature is a combination of original features. For example, principal component analysis [2], local preservation projection [3], the other is based on non-transformed feature selection, also known as band selection. Compared with feature extraction, band selection can retain the physical information of the original data. It is an effective method to reduce the dimension.

AP is a sample-based clustering method. As a new clustering algorithm, it was proposed by Frey and Dueck in 2007 [4]. It does not need to specify the number of clusters in advance, and performs clustering based on the similarity between sample points. Regarding all sample points as the initial clustering exemplar, the AP executes fast and has small errors, and the most representative features can be found in a short time. It has been successfully applied to face recognition [5] and wireless remote [6]. [7] combine AP with ternary mutual information to measure the band correlation of classification. In [8], the relationship between hyperspectral data points is measured by two constraints, and a good classification effect is achieved. AP controls the number of identified clusters by parameters called preferences, but only adjusting preferences cannot find all ideal cluster exemplars.

This paper proposes an improved AP algorithm that divides each intrinsic cluster into several subsets. In theory, this segmentation can be randomly, which can obtain suitable number of clustering results. IAP can find the same number or more ideal band subsets as AP.

2. ALGORITHM PROPOSED

2.1. Affinity propagation

AP is a sample-based clustering method [4]. It selects real data points as exemplars, treats all sample points as initial exemplars, and exchanges information between data points until a representative sample set appears. AP takes the similarity between data points as input. The similarity is defined by the negative Euclidean distance. These

similarities can be symmetrical or asymmetrical. Similarity measure can be calculated as

$$s(x_i, x_j) = -\|x_i - x_j\|^2 \quad (1)$$

$s(x_i, x_j)$ represents the similarity of x_i and x_j in AP which is called preference.

Two types of messages are passed in the AP algorithm, which are responsibility and availability. $r(x_i, x_j)$ represents the numerical message sent from x_i to candidate exemplar x_j , reflecting whether x_j is suitable as the exemplar of point x_i , $a(x_i, x_j)$ represents numeric message sent from candidate exemplar x_j to x_i , reflecting whether x_i chooses x_j as its exemplar. The larger value of $r(x_i, x_j)$, the greater the probability that x_j is the exemplar, and the greater probability that x_i belongs to x_j cluster. AP algorithm continuously updates the attractiveness and attribution value of each point through an iterative process, until a number of high-quality exemplars are generated, and the remaining data points are assigned to the corresponding clusters.

Initially, the availabilities are set to zero, i.e., $a(x_i, x_j) = 0$, updating with the following formula

$$r(x_i, x_j) = s(x_i, x_j) - \max_{k \neq j} [s(x_i, x_k) + a(x_i, x_k)] \quad (2)$$

$$a(x_i, x_j) = \begin{cases} \sum_{k \neq j} \max[0, r(x_k, x_j)] & i = j \\ \min[0, r(x_j, x_i) + \sum_{k \neq j, i} \max[0, r(x_k, x_j)]] & i \neq j \end{cases} \quad (3)$$

After each update, the exemplar x_j of the current sample x_i can be determined, x_j is the exemplar of the maximum value of $a(x_i, x_j) + r(x_i, x_j)$. If $i = j$, then sample x_j is the exemplar of its own cluster, otherwise, x_i belongs to the cluster to which x_k belongs. Finally, all the exemplar points are found by the main diagonal elements of the decision matrix, which meet $a(x_i, x_j) + r(x_i, x_j) > 0$.

In practical applications, search algorithms are prone to shocks. Therefore, a damping coefficient is added to ensure that the algorithm converges

$$\begin{aligned} \mathbf{R}^{t+1} &= \alpha \mathbf{R}^{t+1} + (1-\alpha) \mathbf{R}^t \\ \mathbf{A}^{t+1} &= \alpha \mathbf{A}^{t+1} + (1-\alpha) \mathbf{A}^t \end{aligned} \quad (4)$$

Where \mathbf{R} and \mathbf{A} represent the responsibility matrix and the

availability matrix, t is the number of iterations, α is the damping factor with range of $[0.5, 1]$.

2.2. Improved AP

The information entropy of an image can be used as a quantification standard for the image. The larger the information entropy is, the more information the image contains. The information entropy I of a one-dimensional image can be expressed as

$$I = -\sum_{i=0}^{255} p(i) \log p(i) \quad (5)$$

Where $p(i)$ represents the probability that a pixel with a grayscale value of i appears in the grayscale image.

The initial availability matrix is an important factor leading to different exemplar sets. This paper proposes a new method to change exemplar sets by changing the initial availability matrix.

Algorithm IAP

The preferences $s(x_i, x_j)$ of band x_i and band x_j is calculated according to formula (1). Thus, the similarity matrix $\mathbf{S}_{L \times L}$ is formed, where L represents the total number of bands of hyperspectral image.

Input: $\mathbf{S}_{L \times L}$, number of sub-blocks: k .

Output: exemplar of each data point, i.e., selected band subsets.

Steps:

1. The information entropy of each band image is calculated according to formula (5), then arranged in descending order to \mathbf{ent} , Permute $\mathbf{S}_{L \times L}$ according to \mathbf{ent} and get the permuted $\mathbf{S}'_{L \times L}$, that is:

$$s'(x_i, x_j) = s(\mathbf{ent}(x_i), \mathbf{ent}(x_j)) \quad (6)$$

2. Divide matrix $\mathbf{S}'_{L \times L}$ into k blocks, that is:

$$\mathbf{S}' = \begin{bmatrix} \mathbf{S}'_{11} & \mathbf{S}'_{12} & \cdots & \mathbf{S}'_{1k} \\ \mathbf{S}'_{21} & \mathbf{S}'_{22} & \cdots & \mathbf{S}'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}'_{k1} & \mathbf{S}'_{k2} & \cdots & \mathbf{S}'_{kk} \end{bmatrix} \quad (7)$$

k must be larger than 1 and less than $\sqrt{L}/2$.

3. Use sub-matrix $\mathbf{S}'_{11}, \mathbf{S}'_{22}, \dots, \mathbf{S}'_{kk}$, as input of AP, then we get k availability matrices $\mathbf{A}'_{11}, \mathbf{A}'_{22}, \dots, \mathbf{A}'_{kk}$.

4. Combine $\mathbf{A}'_{11}, \mathbf{A}'_{22}, \dots, \mathbf{A}'_{kk}$ as the input of the initial

availability matrix of AP to get the exemplar set I :

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A}'_{11} & & & \\ & \mathbf{A}'_{22} & & \\ & & \ddots & \\ & & & \mathbf{A}'_{kk} \end{bmatrix} \quad (8)$$

The rest of matrix \mathbf{A}' are all zeros.

5. Restore I :

$$I(x_i) = \text{ent}(I'(x_i)) \quad (9)$$

6. Output exemplar of each data point, i.e., selected band subsets.

3. EXPERIMENTAL RESULTS

3.1. Dataset: Salinas scene

Salinas scene is a public hyperspectral data set acquired by the AVIRIS sensor over Salinas Valley, California, and with a spectral resolution of 3.7m per pixel and the spectral resolution is 10nm. The image size is $512 \times 227 \times 224$, including 20 water absorption bands, 108-112, 154-167, and 224. After removing water absorption bands, there are 204 bands remaining. Fig 1 shows the pseudo-color image of Salinas scene and 16 category distribution labels.

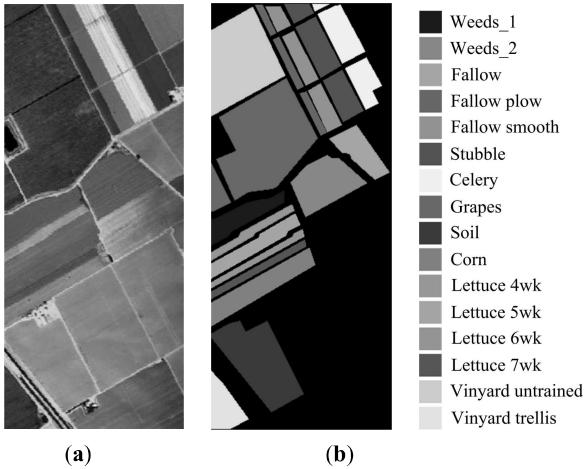


Fig 1. Salinas scene with 16 classes, with (a) pseudo-color of Salinas scene and (b) Color ground-truth image with class labels

The experimental results are verified by hyperspectral classification using support vector machine (SVM) algorithm. Gaussian radial basis kernel function

is selected, and in order to ensure the fairness of experimental results, all algorithms use the same training and testing sets.

Table 1 shows Classification accuracy of all bands, AP, IAP on Salinas scene, where the boldface indicates the optimal classification accuracy. For each method, the average OA, AA, Kappa, the number of selected bands and the corresponding standard deviation in 30 independent runs are recorded. It can be seen that the number of band combinations selected by AP and IAP is fixed, and the experimental results won't be changed with the number of experiments, which also shows the stability of the algorithm. IAP obtains the highest classification accuracies in terms of OA, AA, and kappa indexes.

Table 1. Classification accuracy of all bands, AP, IAP on Salinas scene

Category	All bands	AP	IAP
1	1 ± 0.5	99.93 ± 0.9	99.85 ± 0.4
2	99.87 ± 0.4	99.74 ± 0.7	99.71 ± 0.6
3	99.32 ± 1.7	95.86 ± 1.9	96.09 ± 2.0
4	98.77 ± 0.3	99.32 ± 0.7	98.64 ± 0.2
5	99.60 ± 0.6	99.30 ± 0.9	99.10 ± 0.7
6	99.97 ± 0.9	99.79 ± 1.3	99.97 ± 0.8
7	99.66 ± 0.2	99.86 ± 0.3	99.42 ± 0.7
8	81.40 ± 2.3	81.90 ± 3.9	81.87 ± 2.7
9	99.33 ± 0.7	99.60 ± 1.1	99.51 ± 1.0
10	90.73 ± 1.0	88.03 ± 1.9	94.06 ± 0.8
11	98.52 ± 3.1	89.49 ± 5.4	94.58 ± 4.1
12	96.60 ± 1.7	98.67 ± 3.3	99.45 ± 2.2
13	99.56 ± 1.1	96.19 ± 2.4	99.78 ± 0.9
14	97.25 ± 1.2	93.08 ± 1.9	95.11 ± 1.8
15	74.23 ± 1.1	70.17 ± 1.7	72.62 ± 0.9
16	97.93 ± 1.2	98.52 ± 2.3	98.52 ± 1.7
OA (%)	90.64 ± 1.04	89.34 ± 1.01	91.45 ± 1.03
AA (%)	95.79 ± 1.12	94.34 ± 1.17	95.51 ± 1.09
Kappa (%)	88.15 ± 0.74	88.48 ± 1.80	90.45 ± 0.95
No. band	204 ± 0	7 ± 0	8 ± 0

Fig 2 shows band subsets obtained by AP and IAP, where the white lines indicate the index of the selected band. In order for a visual comparison of classification results, Fig 3 shows classification maps of Salinas image using all bands, band subsets by AP and IAP, respectively. It could be shown from the results that, compared with AP, IAP has a better classification accuracy, which is even better than using total bands.

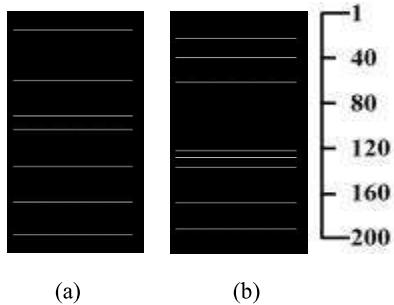


Fig 2. Results of AP and IAP, where white lines represent selected bands, with (a) AP and (b) IAP

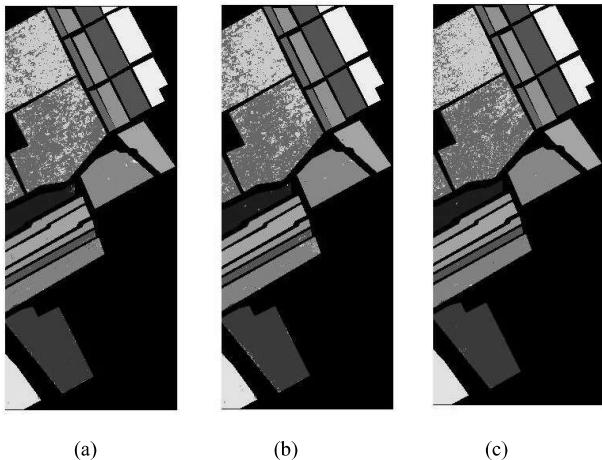


Fig 3. Classification maps of the Salinas image, where (a)-(c) are classification maps by all bands, AP, IAP, respectively.

4. CONCLUSION

This paper proposes an IAP algorithm for hyperspectral classification, which divides each intrinsic cluster into several subsets, and combines the information entropy to change the initial availability matrix to obtain a suitable number of clustering results with arbitrary shapes. The experimental results on the public hyperspectral data set show that the band combination selected by IAP has a better classification effect.

5. ACKNOWLEDGEMENTS

This work is supported by the National Nature Science Foundation of China (61801075), China Postdoctoral Science Foundation (No. 2020M670723), Open Research Funds of State Key Laboratory of Integrated Services Networks, Xidian University (No. ISN20-15) and the Fundamental Research Funds for the Central Universities (3132019341).

6. REFERENCES

- [1] M. Zhao, C.Y. Yu, M.P. Song and C.I. Chang, "A Semantic Feature Extraction Method for Hyperspectral Image Classification Based on Hashing Learning," *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, Netherlands, 2018.
- [2] M.D. Farrell and R. M. Mersereau, "On the Impact of PCA Dimension Reduction for Hyperspectral Detection of Difficult Targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp.192-195, 2005.
- [3] L. Lei, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185-1198, 2012.
- [4] B.F. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *Science*, vol. 315, pp.972- 976, 2007.
- [5] R. Rina, B.M. Achmad, J. Asep and S. Adang, "Clustering Grey-Scale Face-Images Using Modified Adaptive Affinity Propagation with a New Preference Model," *International Conference on Informatics and Computing (ICIC)*, Palembang, Indonesia, 2018.
- [6] S. Park, H.S. Jo, C. Mun and J.G. Yook, "Radio Remote Head Clustering with Affinity Propagation Algorithm in C-RAN," *IEEE Conference on Vehicular Technology (VTC)*, Honolulu, HI, USA, 2019.
- [7] L.C. Jiao, J. Feng, F. Liu, T. Sun and X.G. Zhang, "Semisupervised Affinity Propagation Based on Normalized Trivariable Mutual Information for Hyperspectral Band Selection," *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2760-2773, 2015.
- [8] C. Yang, L. Bruzzone, H.S. Zhao, Y.C. Liang and R.C. Guan, "Decorrelation-Separability Based Affinity Propagation for Semisupervised Clustering of Hyperspectral Images," *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 9, no. 2, pp. 568-582, 2016.