Title: A Feature-Metric-Based Affinity Propagation Technique for Feature Selection in Hyperspectral Image Classification

This paper appears in: Geoscience and Remote Sensing Letters, IEEE

Date of Publication: Sept. 2013

Author(s): Chen Yang; Sicong Liu; Bruzzone, L.; Renchu Guan; Peijun Du

Volume: 10, Issue: 5

Page(s): 1152-1156

DOI: 10.1109/LGRS.2012.2233711

A Feature-Metric-Based Affinity Propagation Technique for Feature Selection in Hyper-spectral Image Classification

Chen Yang, *Member, IEEE*, Sicong Liu, Lorenzo Bruzzone, *Fellow*, *IEEE*,
Renchu Guan, *Member, IEEE* and Peijun Du, *Senior Member*, *IEEE*

***Abstract*—Relevant component analysis (RCA) has shown effective in metric learning. It finds a transformation matrix of the feature space using equivalence constraints. This paper explores the idea for constructing a feature metric (FM) and develops a novel semi-supervised feature selection technique for hyperspectral image classification. Two feature measures referred to as band correlation metric (BCM) and band separability metric (BSM) are derived for the FM. The BCM can measure the spectral correlation among the bands, while the BSM can assess the class discrimination capability of the single band. The proposed feature-metric-based affinity propagation (FM-AP) technique utilize an exemplar-based clustering, i.e. affinity propagation (AP) to group bands from original spectral channels with the FM. Experimental results are conducted on two hyperspectral images and show the advantages of the proposed technique over traditional feature selection methods.**

***Index Terms*—affinity propagation, band selection, feature metric, feature selection, hyperspectral images, relevant component analysis, remote sensing**

I. INTRODUCTION

HYPERSPECTRAL sensors acquire data simultaneously in hundreds of narrow and adjacent spectral channels [1]. This results in high potentialities for detailed land-cover classification but also involves several significant challenges to the classification process, such as: 1) the Hughes phenomenon (decrease of

This work was supported in part by the Erasmus Mundus Action 2 Lot 11 Asia Regional Program, by the National Natural Science Foundation of China (No.41101376, No.61103092), by the China Postdoctoral Science Foundation (No.20110491305), by the Special Fund for Basic Scientific Research of Central Colleges, Jilin University (No.201103036).
C. Yang is with the College of Earth Sciences, Jilin University, Changchun, 130061, China (e-mail: yangc_616@yahoo.com.cn).
S.C. Liu and L. Bruzzone are with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: sliu@disi.unitn.it; lorenzo.bruzzone@ing.unitn.it).
R. C. Guan is with the College of Computer Science and Technology, Jilin University, Changchun, 130012, China (e-mail: guan_renchu@yahoo.com.cn).
P.J. Du is with the Department of Geographical Information Science, Nanjing University, Nanjing, 210000, China (e-mail: dupjrs@126.com).

classification accuracy by increasing the number of spectral channels due to the small ratio between number of training samples and number of spectral channels); and 2) the high computational time required for the analysis of large images [2]. As a result, it is necessary to reduce the number of channels to give as input to a classifier keeping the most informative features and removing redundant channels. Band (feature) selection techniques can be used for identifying a subset of discriminative and uncorrelated features for improving the classification result [3]. In general, a feature selection technique includes two main parts: a criterion function and a search strategy. The former measures the effectiveness of considered subset of features, while the latter is based on an algorithm that explores the space of solutions for finding a subset of features that optimizes the adopted criterion function [4],[5].

The most effective feature selection techniques are those supervised, i.e. techniques that require the availability of a training set. However, true labels of samples (i.e., land-cover class labels) are expensive to achieve and thus it is difficult to define a reliable and complete training set. On the contrary, in many hyperspectral image classification problems, it is quite easy for a user to define pairwise constraints, e.g., indications on the fact that some pairs of samples may belong to the same class or not. Recently, learning distance functions with prior knowledge for semi-supervised learning applications have been studied [6-8]. Among these techniques, relevant component analysis (RCA) [6],[9],[10] is a relatively simple and efficient method for learning the Mahalanobis metrics in semi-supervised fashion. Unlike other methods, it utilizes prior knowledge expressed as equivalence constraints. In other words, it assumes to know that small groups of samples belong to the same class, but without knowing their labels. These small groups of points are termed as “chunklets”.

The search strategy can also be considered as a process of feature clustering, which partitions features into similar groups based on the defined criterion function. An ideal search strategy finds features exhibiting both low correlation and high discrimination ability. Recently, an exemplar-based clustering algorithm, i.e. affinity propagation (AP), was proposed in the literature [11]. It takes as input similarities between data points and finds clusters with small error, especially for large data sets, with fast execution speed. It has been applied to some different fields, e.g. face recognition, gene finding, remote sensing images, text mining [12]-[15].

In this letter, we propose to use the RCA to learn a whitening transformation matrix based on equivalence constraints with the goal of constructing the feature metric (FM). We consider two kinds of metrics, i.e. band correlation metric (BCM) and band separability metric (BSM) in the FM for measuring the spectral correlation and assessing the class discrimination capability. Then, the proposed FM is introduced into the AP, thus defining a novel semi-supervised band selection method, which is called feature-metric-based affinity propagation (FM-AP). The experimental results obtained on two hyperspectral image data sets point out the effectiveness of the proposed method.

II. Proposed Semi-supervised Band Selection Method

Let $X = \{x_1, x_2, \ldots, x_D\} \subset \mathbb{R}^{N\times D}$ a hyperspectral data set, where $x_i = \{x_{i1}, x_{i2}, \ldots, x_{iN}\}$, D is the number of spectral bands, and N is the number of pixel points. Let $H_k = \{x_{k1}, x_{k2}, \ldots, x_{kn_k}\}$, k=1,2, …,K, be K chunklets, where n_k is the number of points in the kth chunklet. The goal of feature selection is to find a subset of uncorrelated bands $Y = \{y_1, y_2, \ldots, y_d\}, y_j \in X$, ($d<D$) that can effectively discriminate the classes present in the data.

A. Feature Metric

In this section we first introduce relevant concepts related to relevant component analysis (RCA). Then the mathematical definition of the feature metric (FM) is given.

In hyperspectral imagery, two pixels x_1 and x_2 can be defined to be related by a positive constraint when they share the same (unknown) label. If pixels x_1 and x_2 are related by a positive constraint and the same hold also for x_2 and x_3, then a chunklet $\{x_1, x_2, x_3\}$ is formed.

Relevant component analysis (RCA) uses a whitening transformation based on the class covariance for rescaling the feature space. This gives the transformation $W = V\Lambda^{-1/2}$, where V and Λ can be found by the singular value decomposition of the class covariance.

1) Chunklet scatter matrix

It is straightforward to estimate the class covariance with labeled data. In RCA, an approximation of the class covariance can be calculated using chunklets assuming that the labeled samples are not available. The chunklet scatter matrix S_{ch} is calculated by [7]:

$$S_{ch} = \frac{1}{|\Omega|} \sum_{k=1}^{K} |H_k| Cov(H_k) \tag{1}$$

where H_k denotes the samples of the kth chunklet, $\cup H_k = \Omega$; $|H_k|$ is the size of the kth chunklet, and K is the number of chunklets. We look for a chunklet that achieves approximation of the mean value of a class, regardless of the size of the chunklet.

2) Whitening transformation

The whitening transformation W can be learned as follows:

(i) Compute the chunklet scatter matrix S_{ch}. Let e denote the effective rank of S_{ch}.

(ii) Calculate the total scatter matrix of the original data, and project the data using principal component analysis (PCA) to its e largest dimensions.

(iii) The chunklet scatter matrix S_{ch} is projected onto a reduced dimensional space and the corresponding whitening transformation W can be computed.

3) Feature metric

The feature metric contains two measures: band correlation metric (BCM) and band separability metric (BSM).

Band correlation metric (BCM)

Band correlation metric (BCM) between two different bands $x_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]$ and $x_j = [x_{j1}, x_{j2}, \ldots, x_{jN}]$ is expressed as:

$$BCM(x_i, x_j) = -|W(x_i, x_j)|^{-1} \tag{2}$$

$$i = 1,2,\cdots,D;\ j = 1,2,\cdots,D;\ i \neq j$$

Band separability metric (BSM)

The band separability metric (BSM) of the band x_p= [x_{p1}, x_{p2},…, x_{pN}] is defined as:

$$BSM(x_p) = -FTS \cdot (W(x_p, x_p)^{-1} \cdot \frac{Max}{Min}) \tag{3}$$

$$p = 1,2,\cdots,D$$

where *Max* and *Min* are the maximum and minimum values of $W(x_p,x_p)$, respectively. *FTS* (called *Feature Threshold Scalar*) is used to get the expected number of bands through setting appropriate values.

The feature metric (FM) is computed based on the BCM and the BSM. By introducing the whitening transformation into the FM we can make use of the equivalence constraints to measure the effectiveness of each band. Thus FM can take into account both the correlation between all pairs of bands and the class discrimination capability of each band.

B. *Feature-Metric-Based Affinity Propagation (FM-AP)*

Affinity propagation (AP) was proposed as a technique for exemplar learning that aims to identify exemplars among data points and forms clusters of data points around these exemplars. Each exemplar is a data point that represents itself and the related cluster of the other data points. AP takes input measures as similarity between pairs of data points. In AP, a common choice for similarity is the negative Euclidean distance, even if more general notions of similarity can be used. The similarities may be positive or negative.

According to the feature metric proposed in Section II.A, the BCM is the similarity between two different bands, and the BSM is the preference, which is the prior suitability to serve as a representative band. Thus, we introduced the developed FM into the affinity propagation (AP) algorithm as the criterion for band selection. The resulting technique is called feature-metric-based affinity propagation (FM-AP) band selection.

AP is derived from factor graph which is constructed by net similarity. Different types of messages that need to be propagated in the factor graph can be reduced to two simple sets of messages that are iteratively updated until convergence. The two kinds of messages are responsibility *r* and availability *a*, and each takes into account a different kind of competition.

The availabilities are initialized to zero, i.e. $a(x_i, x_j) = 0$. Then, the rules of availability and responsibility between two bands x_i and x_j are updated by using the max-product algorithm and are as follows:

$$a(x_i, x_j) = \begin{cases} \sum_{k \neq j} \max\{0, r(x_k, x_j)\} & i = j \\ \min\left\{0, r(x_j, x_j) + \sum_{k \neq j,i} \max\{0, r(x_k, x_j)\}\right\} & i \neq j \end{cases} \tag{4}$$

$$r(x_i, x_j) = s(x_i, x_j) - \max_{k \neq j}\{s(x_i, x_k) + a(x_i, x_k)\} \tag{5}$$

The responsibility $r(x_i,x_j)$, sent from band x_i to representative band x_j, indicates how well-suited the channel x_i would be as a member of the representative band x_j. The "availability" $a(x_i,x_j)$, sent from representative channel x_j to its potential member band x_i, indicates the capability of representative band x_j to represent band x_i. When the searching algorithm converges, a subset of optimal bands is obtained by calculating the set of positive $a(x_i, x_i) + r(x_i, x_i)$ messages for each band x_i. The use of simple updating rules for computing responsibilities and availabilities may result in undesired oscillations. Thus damping is commonly used in over-relaxation methods to avoid numerical oscillations. The two kinds of messages can be damped according to the following equations:

$$\begin{aligned} R^{t+1} &= \alpha R^{t-1} + (1-\alpha) R^t \\ A^{t+1} &= \alpha A^{t-1} + (1-\alpha) A^t \end{aligned} \tag{6}$$

where R and A represent responsibility and availability vectors, respectively; α is the factor of damping (which should satisfy $0.5 \leq \alpha < 1$), and t is the number of iterations. Higher values of α involve slower convergence. The assignment of representative bands is done according to the following rule:

$$\max_{x_j \in Y}\{a(x_i, x_j) + r(x_i, x_j)\} \tag{7}$$

The procedure associated with the proposed FM-AP is as follows.

Step 1 - Set initial values of exemplars and parameters

At beginning, we simultaneously consider all spectral bands to be initial clustering exemplars, i.e., representative bands ($Y = X$). At the same time, we set the chunklet information using the equivalence constraints.

Step 2 - Calculate chunklet scatter matrix and whitening transformation

In practical applications, we need to get the covariance matrix $Cov(H_k)$ for estimating the chunklet scatter matrix, which is based on equivalence constraints only, and does not use any explicit label information. Accordingly, the covariance matrix of chunklet can be obtained as follows:

$$Cov(H_k) = \frac{1}{|H_k|}\sum_{i=1}^{n_k}(x_{ki} - \mu_k)(x_{ki} - \mu_k)^T \qquad (8)$$

$$i = 1,2,\cdots,n_k; k = 1,2,\cdots,K$$

where μ_k is the mean of chunklet H_k, $\mu_k = \frac{1}{|H_k|}\sum_{x \in H_k} x$.

Then the chunked scatter matrix S_{ch} can be obtained as:

$$S_{ch} = \frac{1}{|\Omega|}\sum_{k}^{K}\sum_{i=1}^{n_k}(x_{ki} - \mu_k)(x_{ki} - \mu_k)^T \qquad (9)$$

Afterwards, data set is whitened with respect to the estimated within class covariance matrix. The whitening transformation assigns lower weights to the directions of large variability, as this variability is mainly due to within class changes and is irrelevant to the task of classification. The whitening transformation can be computed from S_{ch} by using the following equation:

$$W = S_{ch}^{-\frac{1}{2}} \qquad (10)$$

Step 3 - Calculate the FM for all spectral bands

FM is computed according to (2) and (3).

Step 4 - Update responsibility and availability

Responsibility and availability are updated according to (4), (5), and (6).

Step 5 - Identify the representative bands and their number

This is done according to (7).

Step 6: Convergence

Repeat Steps 4-5 until the decisions for the representative bands and cluster boundaries are unchanged for some number of iterations.

The subset of bands Y obtained at convergence is the final feature selection result.

III. EXPERIMENTAL RESULTS

A. Data Description

The experimental analysis was done on the two different hyperspectral data sets described as follows.

Hyperion data set

The first data set is a subset of 250×566 pixels of a hyperspectral image acquired by Hyperion on February 7th, 2004 in the urban area of Xuzhou city, Jiangsu Province, China. The original image contains 224 spectral channels with wavelength range from 356 to 2577 nm, where only 198 bands are calibrated. A set of 152 bands was selected for our test after removing the bands which are uncalibrated, with corrupted strips or low image quality. We considered five classes, i.e., water, vegetation, woodland, built-up, and bareland, to characterize this area. The reader can refer to [16] for more details about this data set.

AVIRIS data set

The second data set is made up of a public AVIRIS image, i.e., Indian Pines 92AV3C [17] (145 × 145 pixels and 220 bands) acquired on June, 1992 over the northwest Indiana's Indian Pines. It is accompanied by a reference map, indicating partial ground truth. There are 16 land-over classes available in the original ground truth composed basically of different crop types, vegetation, and man-made structures. In our experiments 200 bands were pre-selected in the data set after discarding the lower signal-to-noise (SNR) bands (104-108, 150-163, 220). As described in [18], only 9 out of 16 classes were considered as the remaining 7 classes were represented by very few labeled samples that do not make it possible a reliable statistical validation.

B. Design of Experiments

In order to assess the effectiveness of the proposed FM-AP technique, it was compared with the following band selection methods: *i*) variance-based band selection (maximum-variance principal component analysis, MV-PCA) [19]; *ii*) clustering-based band selection (standard AP based on the Euclidean distance, ED-AP); and *iii*) uniform band selection (Pearson correlation coefficient, PCC). We also compared the results with that obtained by using all the original bands (Baseline). It is worth noting that in order to have a fair comparison we did not consider supervised feature selection techniques in the experimental analysis.

The performance of each band selection technique was evaluated by using the classification accuracy provided by the widely used support vector machine (SVM) classifier [20]. For comparing the performance of the four algorithms accurately, for Hyperion/AVIRIS data sets, we randomly selected 1980 /3990 samples as equivalence constraints (which are used for both FM-AP learning and SVM training) and 4253/5355 samples as test data (which are used for accuracy assessment) from the available ground truth dataset. In order to have results statistically significant we repeated the process five times and reported the average results. The number of samples used as equivalence constraints and that of samples included in the test set are shown in Table I. The averaged overall accuracy is used to evaluate the results of the band selection.

TABLE I

NUMBER OF PIXELS USED AS EQUIVALENCE CONSTRAINTS / CHUNKLETS AND INCLUDED IN THE TEST SET FOR THE HYPERION AND AVIRIS DATA SETS

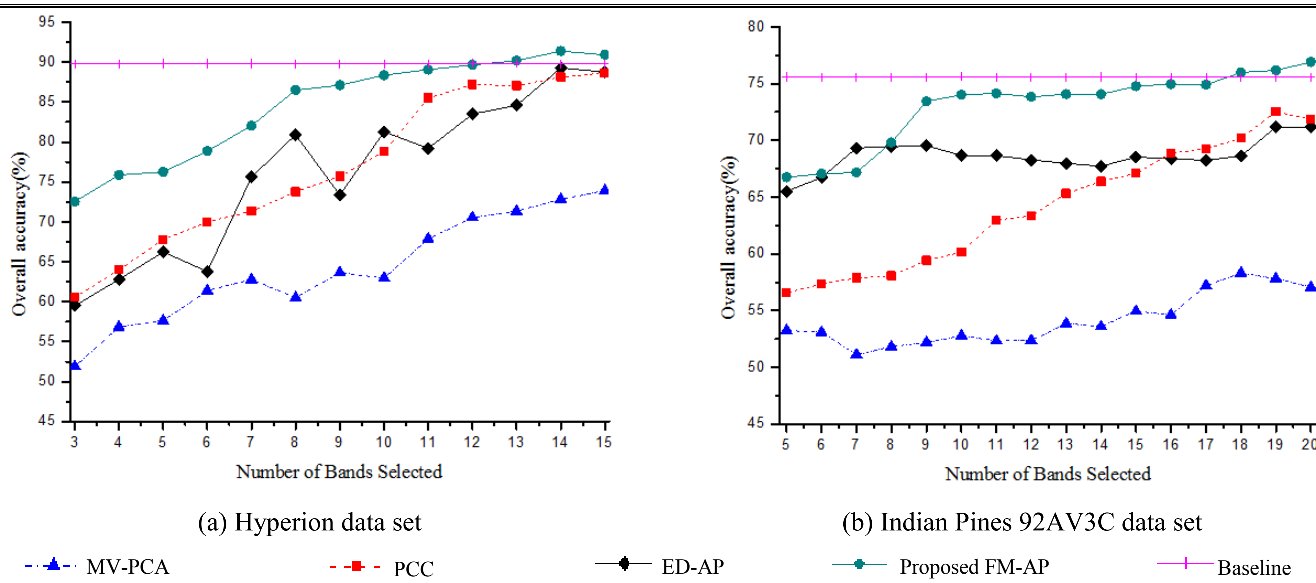
Data set	**Class**	**Equivalence/ Chunklet**	**Test set**	**Total**
Hyperion	Water	360/2	831	1191
	Vegetation	540/3	1110	1650
	Woodland	360/2	797	1157
	Built-up	540/3	1185	1725
	Bareland	180/1	330	510
AVIRIS	Corn-notill	630/3	804	1434
	Corn-min	420/2	414	834
	Grass/Pasture	210/1	287	497
	Grass/Trees	210/1	537	747
	Hay-windrowed	210/1	279	489
	Soybeans-notill	420/2	548	968
	Soybeans-min	1050/5	1418	2468
	Soybean-clean	210/1	404	614
	Woods	630/3	664	1294

C. Results

In this section, comparisons among the results provided by the proposed FM-AP and the MV-PCA, the PCC, the ED-AP, and the Baseline are presented and discussed. Fig.1 (a) and (b) present the averaged overall classification accuracy versus the number of bands selected by using the MV-PCA, the PCC, the ED-AP (α = 0.85) and the FM-AP (α = 0.9) with the Hyperion and AVIRIS data sets, respectively. From Fig.1, one can observe that on the two considered data sets, the classification accuracies basically increased by increasing the number of the selected bands for the four methods. The FM-AP provided the highest accuracy performance compared with the MV-PCA, PCC, and ED-AP algorithms with the same number of selected bands.

In the Hyperion data set experiments (Fig.1(a)), the proposed FM-AP obtained an average overall accuracy of 72.52% when 5 bands were selected. The classification accuracy increased to 86.48% with 8 selected bands. Then, it slowly increased close to 89% with 10 selected channels. It is worth noting that the proposed FM-AP achieved with 13 extracted bands better accuracy (90.16%) than the baseline (89.76%) with all 152 bands. This of course depends on the Hughes phenomenon. The MV-PCA yielded the lowest classification accuracy among the four considered methods. Although ED-AP achieved a few times higher classification accuracies than the PCC, its performance is unstable. On the contrary, the PCC generally achieved a good accuracy, although with slightly lower performance when selecting 7-10 channels.

In the AVIRIS experiments (Fig.1(b)), the two clustering-based methods, i.e. the FM-AP and the ED-AP, obtained higher accuracies than the MVPCA and the PCC. However, the ED-AP decreased the classification accuracy when numbers of selected bands in the range between 9 and 18 were considered. In these conditions, the FM-AP exhibited much more stability. The classification accuracy obtained by selecting 18 channels with the FM-AP (75.99%) was higher than that achieved by using all 200 bands (75.56%). Similarly to the results obtained with the Hyperion data set, the MV-PCA obtained lower classification accuracy than the other three considered methods. The accuracies obtained with the PCC were slightly higher than those obtained with the ED-AP when the number of selected bands was higher than 16. Moreover, it was in general more stable versus the number of selected bands.



(a) Hyperion data set

(b) Indian Pines 92AV3C data set

Fig. 1. Averaged overall accuracy (on five trials) provided by the SVM classifier versus the number of selected bands obtained by the MV-PCA, PCC, ED-AP, and FM-AP methods with (a) Hyperion and (b) AVIRIS data sets. The results achieved by using all the spectral channels are also reported (baseline).

In order to further analyze the effectiveness of the proposed method, we increased the number of selected bands for the four algorithms on the two data sets for comparisons. Tables II presents the results obtained on both considered data sets. It can be seen that the proposed FM-AP technique still performed well in comparison to other approaches when the number of selected bands was further increased. However, as expected, the accuracy decreased by increasing the number of channels over a given value.

TABLE II

COMPARISON OF PROPOSED FM-AP METHOD WITH MV-PCA, PCC, AND ED-AP METHODS BASED ON AVERAGE OVERALL ACCURACY (%) WHEN SELECTED BANDS ARE IN THE RANGE 20-60 FOR BOTH DATA SETS

Data set	Selected bands	MV-PCA	PCC	ED-AP	FM-AP
Hyperion	20	72.82	87.51	82.42	89.23
	30	74.48	88.37	78.35	89.48
	40	75.59	88.83	80.12	90.21
	50	76.06	89.26	85.78	89.83
	60	76.55	89.95	83.27	89.57
AVIRIS	30	58.85	72.62	72.51	77.24
	40	59.27	73.28	73.13	77.67
	50	61.08	74.01	74.02	78.15
	60	62.73	75.62	74.12	78.04

Furthermore, we compared the proposed FM-AP method with the Baseline using different (smaller) number of equivalence constraints on the two data sets. Table III presents the corresponding comparative performance with the average overall accuracy yielded by the FM-AP algorithms on the Hyperion and AVIRIS data sets. For the Hyperion data set, the proposed FM-AP achieved better accuracy (90.43%, 89.83%, and 89.85%) with 19, 37, and 41 selected bands than the baseline (89.76%) when using 1320, 990, and 660 equivalence constraints, respectively.

TABLE III

COMPARISON OF PROPOSED FM-AP METHOD WITH BASELINE BASED ON AVERAGE OVERALL ACCURACY (%) WITH DIFFERENT (SMALLER) NUMBER OF EQUIVALENCE CONSTRAINTS FOR BOTH DATA SETS

Data set	Equivalence constraints	Selected bands	Overall accuracy
Hyperion	1320	19	90.43
	990	37	89.83
	660	41	89.85
AVIRIS	2660	26	75.91
	1995	48	75.62
	1330	45	73.83

For the AVIRIS data set, the classification accuracy obtained by selecting 26 and 48 channels with the FM-AP (75.91% and 75.62%) was higher than that achieved by using all 200 bands (75.56%) when using 2660 and 1995 equivalence constraints, respectively. But when we chose 1330 equivalence constraints, the FM-AP did not increase the accuracy by the number of selected bands was over 41 and even lower than the Baseline. It can be due to the characteristic of the RCA, i.e. learning with limited equivalence constraints may provide unreliable and biased results, especially when the distribution are disequilibrium on the considered data.

D. Parameters Sensitivity Analysis for FM-AP

There are two user-defined parameters in the proposed FM-AP: α, which affects the convergence speed; and *FTS* (Feature Threshold Scalar), which affects the number of selected bands. By increasing the α value the convergence probability increases, but we also increase the execution time (see Table IV). The comparison is performed on PC workstation, (Intel(R) Pentium(R) CPU P600 @ 2.13 GHz, 2.13 GHz with 2.0 GB of RAM).

TABLE IV

EXECUTION TIME (S) VERSUS THE α VALUE FOR THE PROPOSED FM-AP METHOD ON TWO DATA SETS (EQUIVALENCE CONSTRAINTS:1980 /3990 AND SELECTED BANDS: 13/18)

Data set	α	**Execution time (s)**
Hyperion	0.75	78
	0.85	105
	0.95	119
AVIRIS	0.7	83
	0.8	112
	0.9	121

For space constraints, we present here only a sensitivity analysis for *FTS*, as it is related to the quality of the detected subset of bands. We performed experiments by varying the value of *FTS* when running the FM-AP in the two considered data sets. From Fig. 2, one can observe that low values of the *FTS* resulted in the selection of many bands, whereas high values leaded to a small number of bands in all sampling conditions of the two data sets.

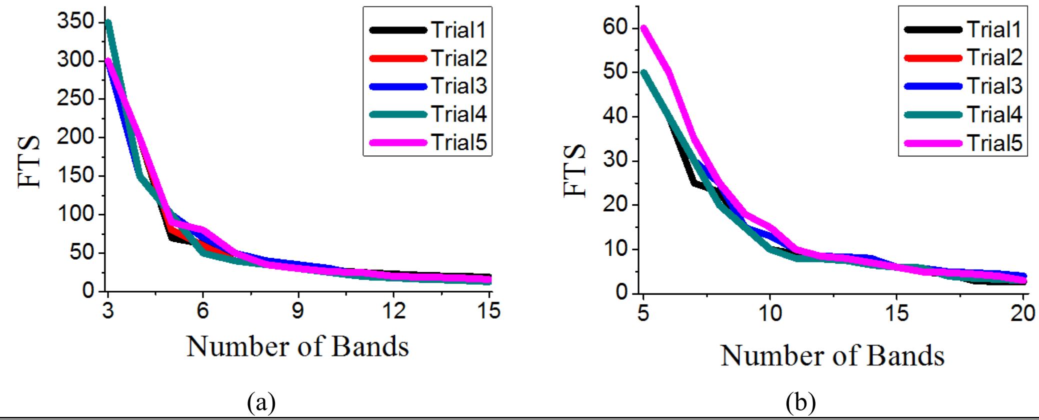


Fig. 2. Number of selected bands versus the value of the FTS parameter for the proposed FM-AP technique on: (a) Hyperion and (b) AVIRIS data sets. Curves with different color describe results obtained in each of the five random experimental trials.

IV. Conclusion

A novel semi-supervised band selection technique, i.e., Feature-Metric-Based Affinity Propagation (FM-AP), has been presented in this paper. The goal of the proposed technique is to make use of equivalence constraints (without assuming availability of class labels) to search a suboptimal feature set for improving the performance of hyperspectral image classification. The proposed FM-AP method takes advantage of the relevant component analysis to build a feature metric (FM) for assessing the class discrimination capability of each band and measuring the spectral correlation between bands. Then, based on the proposed FM, affinity propagation is applied to search a representative subset of spectral channels. Experimental results obtained on two hyperspectral data sets confirm the effectiveness of the proposed FM-AP, which provided high quality feature sets and overcame the Hughes phenomenon in a semi-supervised way.

It is worth noting that the prior information used in the proposed FM-AP method is in the form of positive equivalence constraints. However, negative equivalence constraints may also contain useful information. Therefore, further work is tied to design a method capable to include both positive and negative equivalence constraints in the semi-supervised feature selection.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the free AVIRIS image used in this paper.

REFERENCES

[1] R.B. Smith, *Introduction to hyperspectral imaging*, TNTmips; MicroImages, Inc, 1-24. 2006.

[2] L. Bruzzone, S.B. Serpico, “A technique for feature selection in multiclass problems”, *Inter. J. Remote Sens.*, vol. 21, no. 3, pp. 549-563, 2000.

[3] C.I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Kluwer Academic / Plenum Publishers, New York, 2003.

[4] L. Bruzzone and C. Persello, “A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability,” *IEEE Trans. Geosci. Remote Sens*., vol. 47, no. 9, pp. 3180 - 3191, 2009.

[5] S. B. Serpico and L. Bruzzone, “A new search algorithm for feature selection in hyperspectral remote sensing images,” *IEEE Trans. Geosci. Remote Sens*., vol. 39, no. 7, pp. 1360-1367, 2001.

[6] N. Shental, T. Hertz, D. Weinshall, and M. Pavel, “Adjustment learning and relevant component analysis,” *In Proc.. European Conf. Comput. Vision*, pp. 776-790, 2002.

[7] S. C. H. Hoi, W. Liu, M. R. Lyu, and W. Y. Ma, “Learning distance metrics with contextual constraints for image retrieval,” *In Proc. IEEE Conf. Comput. Vision and Pattern Recogn.*, New York, US, June 17-22 2006.

[8] L. Wu, R. Jin, S.C.H. Hoi, J. Zhu, N. Yu, “Learning bregman distance functions and its application for semi-supervised clustering”, *Adv. Neural Inf. Process. Syst.*, 2009.

[9] A. Bar-Hillel, T. Hertz, N. Shental, D. Weinshall, “Learning distance functions using equivalence relations,” *In Proc. 20th Internat. Conf. Mach. Learn*, Washington, DC, USA, 21-24 August, pp. 11-18, 2003.

[10] A. Bar-Hillel, T. Hertz, N. Shental, D. Weinshall, “Learning a Mahalanobis metric from equivalence constraints,” *J. Mach. Learn. Res*. vol.6, pp.937-965, 2005.

[11] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol.315, pp.972-951, 2007.

[12] B. J. Frey and D. Dueck, “Mixture modeling by affinity propagation,” *in Advances in Neural Inf. Proc. Syst.*, vol. 18. Cambridge, MA: MIT Press. 2, 3, 7, 2006.

[13] B. J. Frey and D. Dueck, “Response to comment on ‘Clustering by passing messages between data points’,” *Science*, vol. 319, no. 5864, p. 726d, Feb. 2008.

[14] C. Yang, L. Bruzzone, F. Sun, L. Lu, R. Guan, and Y. Liang. “A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images.” *IEEE Trans. Geosci. Remote Sens.*, vol.48, no.6, pp. 2647-2659, 2010.

[15] R. Guan, X. Shi, M. Marchese, C. Yang, and Y. Liang. “Text Clustering with Seeds Affinity Propagation,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no.4, pp. 627-637, 2011.

[16] C. Yang, S. C. Liu, L. Bruzzone, R. C. Guan, P. J. Du. “A semi-supervised feature metric based band selection method for hyperspectral image classification,” *4th IEEE GRSS Workshop Hyperspectral Image Signal Process.: evol. remote sens.*, 2012.

[17] AVIRIS NW Indiana’s Indian Pines 1992 data set [Online]. Available: ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C (original files) and ftp://ftp.ecn.purdue.edu/biehl/PC_MultiSpec/ThyFiles.zip (ground truth).

[18] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, 2004.

[19] C.-I. Chang, Q. Du, T. L. Sun, and M. L. G. Althouse, “A joint band prioritization and band-deccorelation approach to band selection for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens*., vol. 37, no. 6, pp. 2631-2641, 1999.

[20] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, U.K.: Cambridge Univ. Press, 2000.