

UNLABELED SELECTED SAMPLES IN FEATURE EXTRACTION FOR CLASSIFICATION OF HYPERSPECTRAL IMAGES WITH LIMITED TRAINING SAMPLES

A. Kianisarkaleh^a, H. Ghassemian^{b,*}, F. Razzazi^a

^a Department of Electrical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran -

azade.kiyani@gmail.com , razzazi@ieee.org

^b Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran - ghassemi@modares.ac.ir

KEY WORDS: Hyperspectral images, feature extraction, limited training samples, unlabeled samples selection, supervised classification.

ABSTRACT:

Feature extraction plays a key role in hyperspectral images classification. Using unlabeled samples, often unlimitedly available, unsupervised and semisupervised feature extraction methods show better performance when limited number of training samples exists. This paper illustrates the importance of selecting appropriate unlabeled samples that used in feature extraction methods. Also proposes a new method for unlabeled samples selection using spectral and spatial information. The proposed method has four parts including: PCA, prior classification, posterior classification and sample selection. As hyperspectral image passes these parts, selected unlabeled samples can be used in arbitrary feature extraction methods. The effectiveness of the proposed unlabeled selected samples in unsupervised and semisupervised feature extraction is demonstrated using two real hyperspectral datasets. Results show that through selecting appropriate unlabeled samples, the proposed method can improve the performance of feature extraction methods and increase classification accuracy.

1. INTRODUCTION

Feature extraction is one of the efficient approaches for overcoming Hughes phenomenon (Hughes, 1968) in hyperspectral image classification. In feature extraction, a transformation is applied by which all pixels in hyperspectral images are transferred from a space with dimension 'd' to a space with dimension 'r' ($r \leq d$) (Jia et al., 2013; Hosseini and Ghassemian, 2015; Imani and Ghassemian, 2015a; Imani and Ghassemian, 2015b). Feature extraction algorithms can be divided into supervised, unsupervised and semisupervised ones. Supervised methods use Labeled samples so they are usually appropriate for classification purposes (Kamandar and Ghassemian, 2013; Imani and Ghassemian, 2014; Imani and Ghassemian, 2015c; Imani and Ghassemian, 2015d). They find a low dimension space where all labeled samples are provided. However supervised FE methods will not be prosperous if a limited number of training samples exist. In this situation unsupervised and semisupervised feature extraction methods show better performance because they can use all pixels in hyperspectral images.

Principal component analysis (PCA) (Fukunaga, 1990) and locality preserving projection (LPP) (He and Niyogi, 2004) are Commonly used unsupervised feature extraction methods. They use all pixels in hyperspectral image without knowing their labels. PCA maintains general information of data by maximizing covariance matrix. LPP preserves local structure of data by forming adjacency graph. semisupervised feature extraction algorithms use unlabeled samples beside labeled samples. Among which are semisupervised Marginal Fisher Analysis (SSMFA) (Huang et al., 2012), semisupervised local discriminant analysis (SELD) (Liao et al., 2013) and semisupervised feature extraction based on supervised and fuzzy-based linear discriminant analysis (SLDA) (Li et al., 2015). These methods maintain general structure of data and increase class separability by using unlabeled and labeled samples.

All unlabeled samples can be applied in unsupervised or semisupervised methods. A hyperspectral image has millions of pixels and using all of them increases storage and calculation costs. So, some unlabeled pixels can be applied in feature extraction process. Chang et al. (2014) defined unlabeled pixels pursuant to the Voronoi cells by using labeled pixels. Shi et al. (2013) determined proper unlabeled pixels using multilevel segmentation results. But unlabeled samples are selected randomly in most semisupervised and unsupervised feature extraction algorithms. Randomly selected samples may include outlier or mixed pixels. Samples may be selected from a limited area. This way, they cannot be an appropriate representative for total data. Therefore, just some unlabeled samples are suitable to be used in feature extraction process. The aim followed by feature extraction in hyperspectral image processing is exploitation of obtained data in classification. So, best result is acquired when data obtained from feature extraction methods that use selected unlabeled samples increase classification accuracy.

This article illustrates the importance of selecting appropriate unlabeled samples used in feature extraction methods. Also it proposes a new method for unlabeled samples selection using spectral and spatial information. Our proposed method has four parts including: PCA, prior classification, posterior classification and sample selection. As hyperspectral image passes these parts, selected unlabeled samples can be used in arbitrary feature extraction methods. In this article, unlabeled selected samples are used in unsupervised LPP and semisupervised SSMFA. We demonstrate that the performance of SSMFA and LPP with selected samples is significantly better than SSMFA and LPP, which use random samples. The experimental results on Pavia University (PU) and Indian Pines (IP) data sets show that appropriate unlabeled samples improves the performance of semisupervised and unsupervised feature extraction methods and increases classification accuracy.

The remaining is organized as followed: next section, details proposed method for unlabeled samples selection. Section 3 overviews LPP and SSMFA methods. Experimental results are reported in the fourth section and last section concludes the paper.

2. THE PROPOSED SELECTION METHOD

A hyperspectral image has millions of pixels and using all of them increases storage and calculation costs. So, some unlabeled pixels can be applied in feature extraction process. We propose a new method for selecting unlabeled samples in this article. The method includes four parts including: PCA, prior classification, posterior classification and sample selection.

2.1 PCA

In the first part, due to limitation in number of training samples and broad dimensions, PCA is used to decrease features. Let $X_{d \times n} = [x_1, x_2, \dots, x_n]$ be the original data and $X_{d \times n_t}^t = [x_1, x_2, \dots, x_{n_t}]$ be the training samples. Where n is the number of total samples, $n_t = \sum_{i=1}^q n_i$ is the number of training samples, n_i refer to the number of training samples in the i th class, q is the number of classes and d denotes the number of spectral bands. $Y_{r \times n} = T^{PCA} \times X$ and $Y_{r \times n_t}^t = T^{PCA} \times X^t$ are projected data using PCA transformation matrix (T^{PCA}). r is the dimensionality of the projected data ($r < d$). Y and Y^t are used in prior classification part. In the experiments, the number of r is empirically set $r = 5$ for $n_t = 10$ and $r = 6$ for $n_t = 15$.

2.2 Prior classification

Prior classification is carried out using Gaussian maximum likelihood (GML), support vector machine (SVM), and K nearest neighbor (KNN) classifiers (Duda et al., 2001; Pal and Mather, 2005). Classification results are obtained in the form of classification map. In classification map, a label ($L_i = 1, \dots, q$) is allocated to each pixel x_i .

The number of neighbors $K = 4$ for KNN classifier is chosen by experiments. Radial basis kernel function (RBF) is used for SVM classifier defined in LIBSVM (Chang and Lin, 2008). The five-fold cross validation is used to choose the optimal penalty parameters C and kernel parameter g in SVM.

2.3 Posterior classification

Posterior classification includes two steps to determine final label of each pixel. Firstly, primary label of each pixel is specified using classification maps. Label L_i will be allocated to x_i if it holds the label in all three classification maps. Otherwise, a zero label is considered for the pixel. In second step, final label of each pixel is obtained using labels in its spatial neighborhood. If, at least, five pixels in 8-spatial neighbors of central pixel x_i contain label L_i , this label will be kept for x_i , otherwise, its label will change to zero.

The objective of second and third sections is to put mixed pixels aside using spectral and spatial information. Mixed pixels are consist of more than one type of land covering and exist because of limitation in spatial resolution. They often found in spatial border of classes. Mixed pixels cannot be perfect representative for classes. So they affect the efficiency of algorithms used to hyperspectral images feature extraction.

2.4 Samples selection

In final part, unlabeled samples are selected among pixels nearest to the mean of classes resulting in the absence of outlier samples in unlabeled ones. Now, we can use the unlabeled selected samples in arbitrary feature extraction. In this article, we use the unlabeled selected samples in unsupervised LPP and semisupervised SSMFA.

3. LPP AND SSMFA

3.1 LPP

LPP is an effective unsupervised method to reduce features (He and Niyogi, 2004). Using LPP, local structure of data is preserved and adjacent samples will be similar in original and transferred space. In order for this, information of adjacent samples is used to form adjacency graph. Let $X_{d \times n_u}^u = [x_1, x_2, \dots, x_{n_u}]$ denote the unlabeled selected samples. Where n_u is the number of unlabeled samples. Each pair of unlabeled samples (x_i, x_j) are considered as nodes for the adjacency graph and an edge is added between x_i and x_j if they are among k nearest neighbors of each other. The weight matrix (A) of adjacency graph is defined as:

$$a_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right) \quad (1)$$

Where t is local scaling parameter. In this article $t = t_i \times t_j$ (Manor and Perona, 2005). Where $t_i = \|x_i - x_i^{(k)}\|$ is a local scaling for x_i and $x_i^{(k)}$ denotes k th nearest neighbors of x_i . Parameters k is not fixed and it is determined using experiments. In this article k is empirically set to 7. LPP transformation matrix (T) is calculated as follows:

$$T = \arg \min_{T \in \mathbb{R}^{d \times r}} (T^* S^L T) \quad (2)$$

subject to $T^* S^N T = I_r$

Where T^* is transpose of T , $S^L = X^u L X^{u*}$, $S^N = X^u D X^{u*}$, D is diagonal matrix with $D_{ii} = \sum_{j=1}^{n_u} a_{ij}$ and $L = D - A$ is laplacian matrix. The solution is given by solving a generalized eigenvalue problem.

3.2 SSMFA

SSMFA is a semisupervised feature reduction method (Huang et al., 2012). In SSMFA geometric structure of labeled and unlabeled data is preserved and labeled data from variable classes are differentiated. In order for this, two between-class (G_b) and within-class (G_w) graphs are used. Let $X_{d \times (n_t + n_u)}$ be the total samples that first n_t samples are labeled and the rest n_u samples are unlabeled. In G_b , each pair of labeled samples (x_i, x_j) is considered as nodes for the graph and an edge is added between x_i and x_j if they have different class labels. In G_w , each pair of samples (x_i, x_j) are considered as nodes for the graph and an edge is added between x_i and x_j if they have the same class labels or at least one of them is unlabeled and they are among k^+ nearest neighbors of each other. The weight matrices W^b of G_b and W^w of G_w is defined as follows:

$$w_{ij}^b = \begin{cases} a_{ij} & \text{if } L_i \neq L_j \\ 0 & \text{if } L_i = L_j \end{cases} \quad (3)$$

$$w_{ij}^w = \begin{cases} \beta a_{ij} & \text{if } L_i = L_j \\ a_{ij} & \text{if } x_i \in N_{k^+}(x_j) \text{ or } x_j \in N_{k^+}(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Where $N_{k^+}(x_i) = \{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(k^+)}\}$ is the k^+ nearest neighbors for x_i , L_i be the x_i label and trade-off parameter $\beta > 1$ adjust the share of unlabeled and labeled data. $a_{ij} = \exp(-\|x_i - x_j\|^2 / \sigma)$ if x_i and x_j are among k nearest neighbors of each other. a_{ij} is calculated by all samples in X . σ is the local scaling parameter. In the experiments parameter β is empirically set to 10. The number of neighbors is set to 5 for $n_i = 10$ and 7 for $n_i = 15$. σ is defined as t in LPP. SSMFA transformation matrix (T) is calculated as follows:

$$T = \arg \max_{T \in \mathbb{R}^{d \times r}} \left(\frac{T^* S^b T}{T^* S^w T} \right) \quad (5)$$

Where $S^b = XL^b X^*$, $S^w = XL^w X^*$, $L^b = D^b - W^b$, $L^w = D^w - W^w$ and D is diagonal matrix with $D_{ii} = \sum_{j=1}^{n_u+n_t} w_{ij}$. The solution is given by solving a generalized eigenvalue problem.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of proposed method, unlabeled selected samples are used in LPP and SSMFA. LPP and SSMFA performance using unlabeled selected samples (LPPUSS and SSMFAUSS) compared to LPP and SSMFA using random selected samples (LPPURS and SSMFAURS), is demonstrated using IP and PU data sets.

IP is related to an agricultural-woodsy area obtained through Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor from a test site in northeast of Indian Pine state. The data includes 224 spectral bands where after 24 noisy bands are removed, experiments are carried out on 200 bands. Each band consists of 145×145 pixels and it includes 21025 samples in total. The ground truth is designated into 16 classes. PU is related to Pavia University area in north Italy obtained by Reflective Optics System Imaging Spectrometer (ROSIS) sensor. It includes 115 spectral bands where after 12 noisy bands are removed, experiments are carried out on 103 bands. Pavia University is 610×340 pixels image and it includes 207700 samples in total. The ground truth differentiates 9 classes.

SVM and GML are used to classify hyperspectral images. RBF kernel is used for SVM classifier defined in LIBSVM. The five-fold cross validation is used to choose the optimal parameters in SVM. The penalty parameter C is tested between $\{10^0, 10^1, \dots, 10^4\}$, and the g parameter is tested between $\{10^0, 10^{-1}, \dots, 10^{-8}\}$. Classification results contain accuracy (Acc) and reliability (Rel) of classes, average accuracy, average reliability, overall accuracy and kappa coefficient (Cohen, 1960). $Acc = \alpha / \beta \times 100$ and $Rel = \alpha / \gamma \times 100$ where α is the number of samples classified correctly in related class, β is the total number of samples belonging to related class and γ is the number of samples classified correctly in related class. Overall accuracy is the percentage of correctly classified samples.

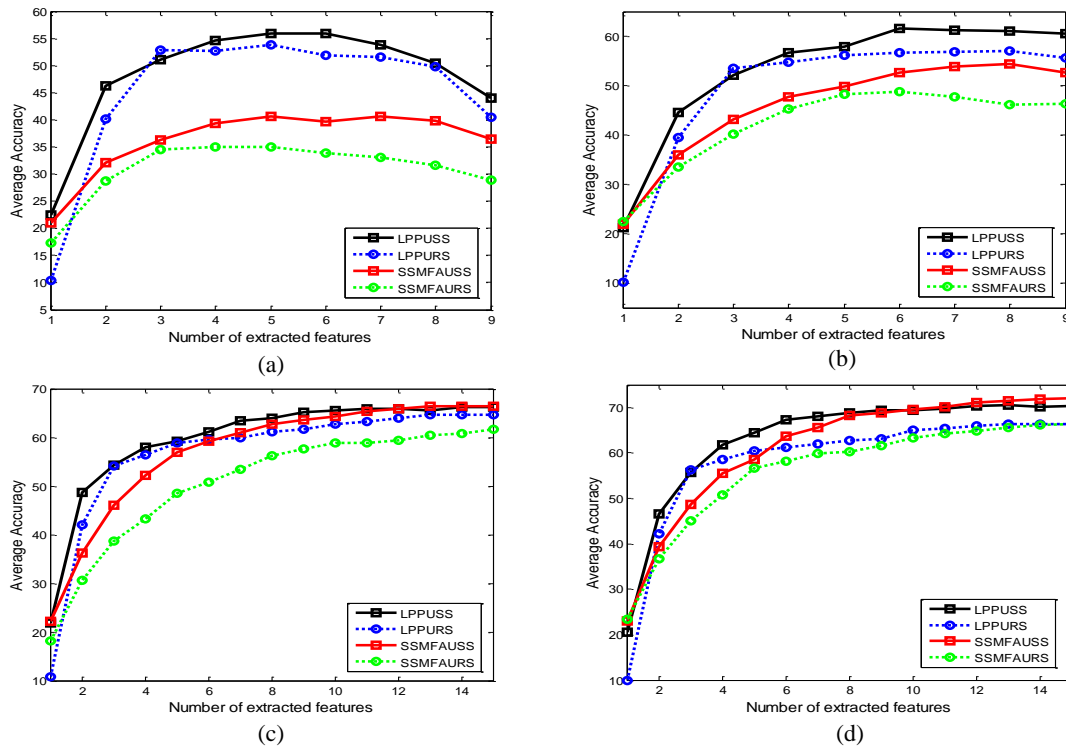


Figure 1. Average accuracy versus the number of extracted features for IP dataset, (a) GML classifier, 10 training samples, (b) GML classifier, 15 training samples (c) SVM classifier, 10 training samples and (d) SVM classifier, 15 training samples.

In order to examine performance of proposed method in classification problems, two experiments are carried out. In first experiment, we evaluate performance of proposed method in ill-posed and poorly-posed classification problems for two datasets. 1400 unlabeled samples and 10 and 15 training samples are used from each class in ill-posed and poorly-posed classification problems respectively. Training samples and random unlabeled samples are chosen 10 times and the average results are reported. Average accuracy (%) versus number of extracted features is shown in figure 1 for IP dataset. Maximum

average accuracy for (a) GML classifier, 10 training samples, (b) GML classifier, 15 training samples (c) SVM classifier, 10 training samples and (d) SVM classifier, 15 training samples are obtained by 5, 6, 15 and 15 extracted features respectively. Figure 2 displays average accuracy (%) versus number of extracted features for PU dataset. Maximum average accuracy for (a) GML classifier, 10 training samples, (b) GML classifier, 15 training samples (c) SVM classifier, 10 training samples and (d) SVM classifier, 15 training samples are obtained by 5, 5, 15 and 15 extracted features respectively.

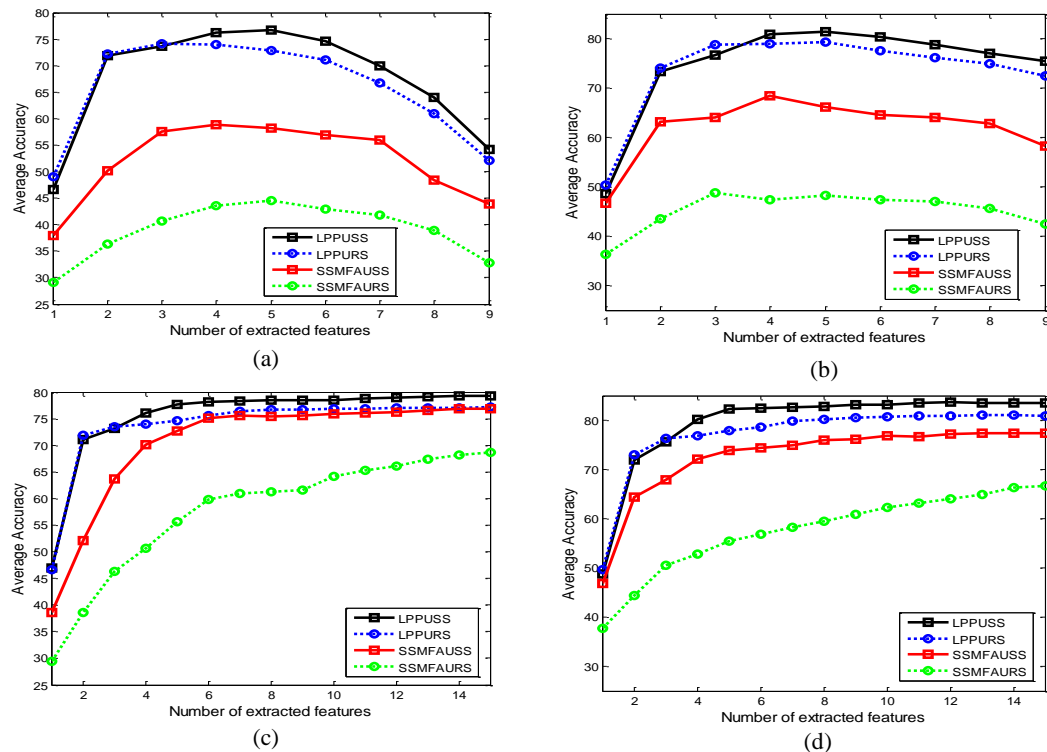


Figure 2. Average accuracy versus the number of extracted features for PU dataset, (a) GML classifier, 10 training samples, (b) GML classifier, 15 training samples (c) SVM classifier, 10 training samples and (d) SVM classifier, 15 training samples

Class		LPPUSS		LPPURS		SSMFAUSS		SSMFAURS	
Name of class	# Samples	Acc	Rel	Acc	Rel	Acc	Rel	Acc	Rel
Asphalt	6631	66.14	94.48	59.84	96.66	60.68	91.53	41.41	77.91
Meadows	18649	70.67	85.54	63.68	86.43	69.05	88.82	52.08	88.23
Gravel	2099	75.37	53.77	61.3	45.92	63.41	52.75	52.8	41.94
Trees	3064	77.22	66.25	82.57	60.44	84.03	67.36	85.49	60.82
Painted metals sheets	1345	99.39	97.84	99.18	96.36	99.23	99.97	94.99	98.82
Bare Soil	5029	60.26	46.85	62.43	41.78	62.78	46.48	66.53	35.79
Bitumen	1330	88.47	47.42	91.67	41.75	82.14	37.31	66.35	23.7
Self-Blocking Bricks	3682	75.75	66.5	73.52	64.22	70.78	54.75	58.62	49.65
Shadows	947	99.93	99.85	99.83	99.83	99.83	96.98	99.77	83.29
Average Acc and Average Rel		79.25	73.17	77.11	70.38	76.88	70.66	68.67	62.24
Overall Acc		71.99		67.81		70		57.97	
Kappa coefficient		64.5		59.82		62.28		49.08	

Table 1. Classification results for PU dataset obtained by SVM classifier, 10 training samples, 1400 unlabeled samples and 15 extracted features

Class		LPPUSS		LPPURS		SSMFAUSS		SSMFAURS	
Name of class	# Samples	Acc	Rel	Acc	Rel	Acc	Rel	Acc	Rel
Alfalfa	46	82.61	31.07	88.26	29.39	85.65	38.11	82.17	33.37
Corn-notill	1428	41.19	45	36.08	43.7	48.68	46.15	41.68	42.5
Corn-mintill	830	40.43	39.04	40.34	37.33	43.64	36.12	41.35	26.9
Corn	237	60.59	29.13	60.25	28.98	53.33	25.22	47.51	26.48
Grass-pasture	483	70.35	60	68.99	48.49	73.04	61.63	65.38	52.68
Grass-trees	730	79.89	82.49	77.29	78.49	81.29	83.6	75.15	77.4
Grass-pasture-moved	28	94.29	43.2	92.86	27.99	92.86	42.32	93.57	26.24
Hay-windrowed	478	70.92	94.62	72.13	94.11	78.74	96.54	74.98	86.67
Oats	20	91	18.86	94	17.55	92	28.78	92	23.06
Soybean-notill	972	47.24	42.73	48.64	42	42.8	39.06	35.58	33.61
Soybean-mintill	2455	45.59	64.97	43.95	63.56	43.34	63.82	38.17	54.56
Soybean-clean	593	41.59	25.79	40.67	24.59	37.27	37.58	24.28	33.97
Wheat	205	95.9	86	90.34	71.15	96.59	75.2	89.66	62.98
Woods	1265	69.11	90.52	62.1	89.47	74.25	90.74	70.99	90.16
Buildings-Grass-Trees-Drives	386	42.75	34.8	29.12	28.98	42.9	42.03	34.35	38.96
Stone-Steel-Towers	93	85.16	83.1	88.6	97.26	77.2	92.95	80.43	90.04
Average Acc and Average Rel		66.16	54.46	64.6	51.44	66.48	56.24	61.7	49.97
Overall Acc		54.17		51.51		55.28		49.49	
Kappa coefficient		48.8		45.92		49.91		43.42	

Table 2. Classification results for IP dataset obtained by SVM classifier, 10 training samples, 1400 unlabeled samples and 15 extracted features

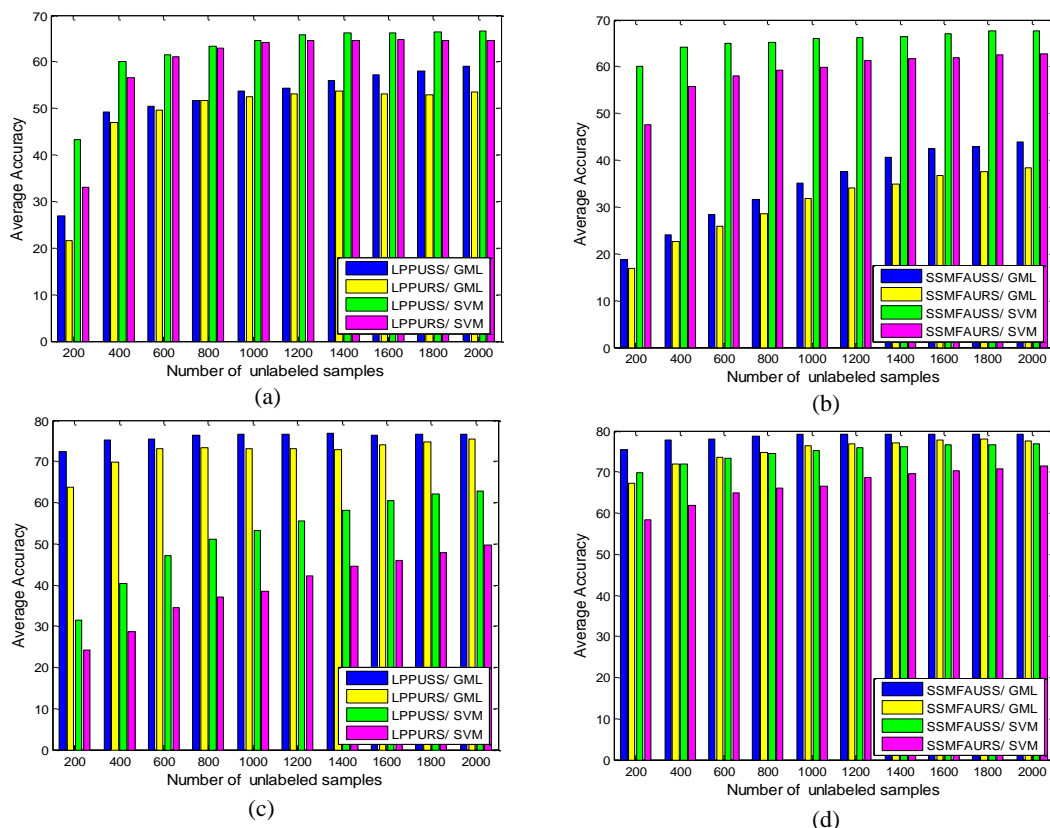


Figure 3. Average accuracy versus the number of unlabeled samples by 10 training samples and 5 and 15 extracted features for GML and SVM classifiers respectively (a) IP, LPP, (b) IP, SSMFA, (c) PU, LPP, (c) PU, SSMFA

Classification results in ill-posed classification problem using SVM classifier and 15 extracted features are reported in Table 1 for PU and Table 2 for IP datasets. In second experiment, number of training samples and extracted features are fixed and number of unlabeled samples varies from 200 to 2000 with a step size increment of 200. Figure 3 shows Average accuracy

versus the number of unlabeled samples for datasets using 10 training samples, 5 and 15 extracted features by GML and SVM classifiers respectively. For both data sets, LPPUSS and SSMFAUSS have better performance compared to LPPURS and SSMFAURS in poorly-posed and ill-posed situations. For all methods, the average accuracy improves as unlabeled

samples increased in number. For IP dataset, maximum average accuracy in ill-posed classification problem with GML classifier using LPPURS and SSMFAURS features are 53.45% and 38.46% respectively. These values are obtained using 2000 unlabeled samples. But obtained average accuracy in the same classification problem with similar classifier are 53.77% using LPPUSS by 1000 unlabeled samples and 40.59% using SSMFAUSS by 1400 unlabeled samples. Similar results are acquired for datasets with GML and SVM classifiers. It can be derived from the results that unlabeled selected samples are an appropriate representative for total data because proposed selection method put mixed and outlier pixels aside using spectral and spatial information. According to experiments, LPPUSS and SSMFAUSS outperform LPPURS and SSMFAURS and they decrease storage and calculation costs using less number of unlabeled pixels. Results show that proposed unlabeled samples improve feature extraction methods performance and increase classification accuracy with limited training samples.

5. CONCLUSION

Selecting appropriate unlabeled samples has important role in the performance of semisupervised and unsupervised feature extraction methods. In this article, using pixels' spatial and spectral information, a method is provided for unlabeled samples selection. Proposed method can solve problems resulted from random selection like outlier or mixed samples. Unlabeled selected samples can be use in any arbitrary feature extraction method. In this article, we use the unlabeled selected samples in LPP and SSMFA. Results show that LPP and SSMFA using unlabeled selected samples provide remarkable results in ill-posed and poorly-posed classification situations.

REFERENCES

- Chang, C., Linin, C., 2008. LIBSVM—A Library for Support Vector Machines. <<http://www.csie.ntu.edu.tw/~cjlin/libsvm>>.
- Chang, Ch. Sh., Chen, K. Ch., Kuo, B. Ch., Wang, M. Sh., Li, Ch. H., 2014. Semi-supervised local discriminant analysis with nearest neighbours for hyperspectral image classification. *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, Quebec, 13-18 July, pp. 1709-1712.
- Cohen, J., 1960. A Coefficient of Agreement From Nominal Scales. *Educ. Psychol. Meas.*, 20(1), pp. 37-46.
- Duda, R. O., Hart, P. E., Stock, D. G., 2001. *Pattern classification*, second ed. John Wiley & Sons Inc., New York.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, Academic Press Inc., San Diego.
- He, X., Niyogi, P., 2004. Locality preserving projections. *Adv. Neural Inf. Process. Syst.*, 16, pp. 153-160.
- Hosseini, S. A., Ghassemian, H., 2015. Application of curve fitting in hyperspectral data classification and compression, *Journal of Information Systems and Telecommunication*, 3(11), pp. 142-150.
- Hughes, G. F., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory*, 14(1), pp. 55-63.
- Huang, H., Liu, J., Pan, Y., 2012. Semisupervised marginal fisher analysis for hyperspectral image classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Melbourne, Australia, Vol. I-3, pp. 377-382.
- Imani, M., Ghassemian, H., 2014. Feature extraction using attraction points for classification of hyperspectral images in small sample size situation. *IEEE Geosci. Remote Sens. Lett.*, 11(11), pp. 1986-1990.
- Imani, M., Ghassemian, H., 2015a. Feature extraction using weighted training samples. *IEEE Geosci. Remote Sens. Lett.*, 12(7), pp. 1387-1386.
- Imani, M., Ghassemian, H., 2015b. Feature extraction using median-mean and feature line embedding. *Int. J. Remote Sens.*, 36(17), pp. 4297-4314.
- Imani, M., Ghassemian, H., 2015c. Feature space discriminant analysis for hyperspectral data feature reduction. *ISPRS J. Photogramm. Remote Sens.*, 102, pp. 1-13.
- Imani, M., Ghassemian, H., 2015d. High-dimensional image data feature extraction by double discriminant embedding, *Pattern Anal. Appl.*, 29, pp. 1-12.
- Jia, X., Kuo, B. C., Crawford, M., 2013. Feature mining for hyperspectral image classification. *Proc. IEEE*, 101(3), pp. 676-697.
- Kamandar, M., Ghassemian, H., 2013. Linear feature extraction for hyperspectral images based on information theoretic learning. *IEEE Geosci. Remote Sens. Lett.*, 10(4), pp. 702-706.
- Li, Ch., Ho, H., Kuo, B. C., Taur, J., Chu, H., Wang, M., 2015. A semi-supervised feature extraction based on supervised and fuzzy-based linear discriminant analysis for hyperspectral image classification. *Appl. Math. Inf. Sci.*, 9(1L), pp. 81-85.
- Liao, W., Pizurica, A., Scheunders, P., Philips, W., Pi, Y., 2013. Semisupervised local discriminant analysis for feature extraction in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* 51(1), pp. 184-198.
- Manor, L. Z., Perona, P., 2005. Self-tuning spectral clustering. *Adv. Neural Inf. Process. Syst.*, 17, pp. 1601-1608.
- Pal, M., Mather, M., 2005. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.*, 26, pp. 1007-1011.
- Shi, Q., Zhang, L., Du, B., 2013. Semisupervised discriminative locally enhanced alignment for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 51(9), pp. 4800-4815.