

結合光譜與空間資訊之線性區別分析應用於 高光譜影像分類

Linear Discriminate Analysis based on both Spectral and Spatial Information for Hyperspectral Image Classification

朱慧珊

國立臺中教育大學
教育測驗統計研究所
碩士生

Roxanne90@gmail.com

李政軒

國立交通大學 電控工程研究所；
國立臺中教育大學
教育測驗統計研究所
博士生

chli.ntcu@gmail.com

郭伯臣

國立臺中教育大學
教育測驗統計研究所
教授兼所長

kbc@mail.ntcu.edu.tw

摘要

在高光譜遙測影像辨識中，經常會遭遇 Hughes 現象，特別容易發生在小樣本的情況。線性區別分析(LDA)是一種常見用來解決 Hughes 現象的特徵萃取(FE)技術。近年來的研究顯示融入空間資訊之演算法皆能大幅提升高光譜影像辨識的結果，故進行高光譜影像辨識時，不能只考慮光譜資訊而忽略空間資訊的重要性。因此，本研究將在 LDA 中融入空間資訊，並採用 Washington DC Mall、Indian Pine Site 等高光譜遙測影像及搭配不同的分類器來進行此分類方式的效能驗證，實驗結果顯示，在有限的小樣本下，可以得到較為有效的分類成果。

關鍵詞：高光譜影像分類、特徵萃取、線性區別分析、光譜與空間資訊

Abstract

Linear discriminate analysis (LDA) is a commonly used feature extract (FE) method to

resolve the Hughes phenomenon for hyperspectral image classification. The Hughes phenomenon is often encountered in hyperspectral image classification, especially in the small sampling size problem. Recent studies show that the spatial information can greatly improve the classification performance of hyperspectral image classification. Hence, for hyperspectral image classification, it is not only necessary to use the available spectral information but also to exploit the spatial information. In this paper, a LDA based on both spectral and spatial information is proposed. The experimental results on the Washington DC Mall and the Indian Pine Site show that the proposed method can yield a better classification performance than LDA in the small sampling size problem.

Keywords : hyperspectral image classification, feature extraction, linear discriminate analysis, spectral and spatial information

1.前言

隨著科技迅速發展，遙測影像資料的可用頻率已推進到現今數百個波段的高維度(Hyper-dimension)資料，維度數的提升對資料型態的描述更為精確，然而由於樣本採集不易，因此造成小樣本、高維度的問題，亦即 Hughes phenomenon [6]的產生，使得辨識效果不佳。

由於訓練樣本的取得不易，且往往需要大量的成本，在解決訓練樣本數不足的問題時，不外乎三種方法：一是設法降低維度(dimensionality Reduction) [2]，另一是正規化(regularization) [3]，三是想辦法增加樣本數的半監督式演算法(semi-supervised algorithm) [1][5][7][8][11][13]。

在傳統辨識方法上通常使用特徵選取(feature selection)或特徵萃取(feature extraction)來降低維度數，而線性區別分析(LDA)是一種常見的參數型特徵萃取技術，且已廣泛應用於高光譜影像辨識及人臉辨識中，故如何在樣本少的情況下，建構出辨識率高的辨識器一直是被關注的研究主題。

近年來研究顯示融入空間資訊之演算法[9][10][12][14][15]皆能有效提升高光譜影像辨識的分類結果，故本研究將在線性區別分析中融入空間資訊，並採用Washington DC Mall、Indian Pine Site等高光譜遙測影像來進行此分類方式的效能驗證，並以數個單步驟辨識器作為比較的基準。

2.方法論

在科技日益發展的今日，高光譜影像資料的維度也隨之增加，為了避免因維度數的增加而造成辨識不佳的情況，採用特徵萃取的技術來改善此問題，是一個值得探究的議題。

在各種特徵萃取技術中，LDA是一個具有簡單易懂及輕便可解性的優勢，所以，若以此

為基礎來進行特徵萃取演算法之改良，會是一個較佳的選擇。以下即針對此特徵萃取的方法來介紹並定義相關的使用符號及數學公式。

L : the number of the classes

N_i : the number of samples in class i ,

$i=1,\dots,L$

$N = N_1 + \dots + N_L$: the number of all training samples

$x_\ell^{(i)}$: the ℓ sample in class i

$H_i = \{x_1^{(i)}, \dots, x_{N_i}^{(i)}\} \subset R^d$: the set of samples in class i

$v^T x_\ell^{(i)}$: sample $x_\ell^{(i)}$ 投影至 v

$A = [v_1, v_2, \dots, v_p]$: the transformation matrix from the original space to the reduced subspace

線性區別分析(Linear Discriminate Analysis, LDA)

Fukunaga(1990)提出線性區別分析(Linear Discriminate Analysis, LDA) [4]，由於需使用樣本平均數與共變異數矩陣，故被視為一種參數型特徵萃取方法。其主要目的為找尋一轉換矩陣 A ，使得資料經轉換後具有最大類別分散量(class separability)，利用組間分散矩陣(S_b)及組內分散矩陣(S_w)之比率來估計類別分散量的準則。

$m_i = \frac{1}{N_i} \sum_{\ell=1}^{N_i} x_\ell^{(i)}$: the mean of the class i

$\bar{m}_i = \frac{1}{N_i} \sum_{\ell=1}^{N_i} v^T x_\ell^{(i)} = v^T \left(\frac{1}{N_i} \sum_{\ell=1}^{N_i} x_\ell^{(i)} \right) = v^T m_i$

: the mean of the class i 投影至 v

$m = \sum_{i=1}^L \frac{N_i}{N} m_i = \sum_{i=1}^L \frac{N_i}{N} \sum_{\ell=1}^{N_i} \frac{1}{N_i} x_\ell^{(i)} = \sum_{i=1}^L \sum_{\ell=1}^{N_i} \frac{1}{N} x_\ell^{(i)}$

: the total mean

組間分散矩陣是以各組期望向量與所有訓練樣本平均之分散程度表示之。

可用方程式表達如下：

$$S_b^{LDA} = \sum_{i=1}^L \frac{N_i}{N} (m_i - m)(m_i - m)^T \quad (1)$$

組內分散矩陣為各類別訓練樣本與其各自的期望向量之分散程度。

可用方程式表達如下：

$$S_w^{LDA} = \sum_{i=1}^L \sum_{\ell=1}^{N_i} \frac{1}{N} (x_{\ell}^{(i)} - m_i)(x_{\ell}^{(i)} - m_i)^T \quad (2)$$

LDA的用途是在取得特徵空間下的最大分散量，故希望組內分散量最小而組間分散量最大，其最佳化下列準則可得到最佳的特徵(optimal features)：

$$J_{LDA} = \text{trace}((S_w^{LDA})^{-1} S_b^{LDA}) \quad (3)$$

亦即求解廣義特徵值問題

$$S_b v_s = \lambda_s S_w v_s, \quad s=1, \dots, d \quad \text{with } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

$$A = [v_1, v_2, \dots, v_p] \quad p \leq d$$

3.光譜與空間資訊之線性區別分析

本研究有鑑於初始訓練樣本取得不易，故藉由空間資訊的概念來增加樣本數，以得到半監督式的訓練樣本，並經由迭代來更新樣本的類別，所提出的方法將可解決樣本數不足的問題及增加分類的辨識率。

3.1 鄰居系統(Neighbor System)

許多研究顯示出空間資訊之間具有相關性，故對影像分類是有助益的，而其中格點(lattice)空間模型更可以描述資料點之間相對應的空間關係，藉此可增加樣本的辨識效果，而Markov Random Field (MRF)[12][15]模型對格點圖像問題更是被常拿來運用，且為一有效分類的方法。故本研究將使用MRF中Neighbor System的概念，以spatial information 來搭配spectral information 所形成的特徵萃取法。

假設一空間資料為S，正規化格點資訊的鄰居系統定義如下所述：

$$\partial S = \{t \in S \mid \text{dist}(s, t) \leq c^2, t \neq s\}$$

其中c代表c階鄰居系統(c-order neighbor system)，其表示如下圖所示：

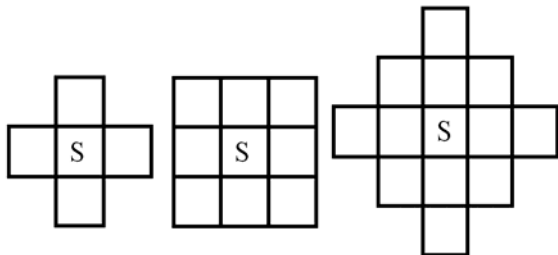


圖 1 由左到右分別為一階鄰居集、二階鄰居集及三階鄰居集

3.2 融入空間資訊之分散矩陣

依據鄰居系統(Neighbor System)的概念，將此運用於線性區別分析中，從而定義新的組間分散矩陣(S_b)及組內分散矩陣(S_w)，並以參數組合來調整光譜及空間資訊的重要性。

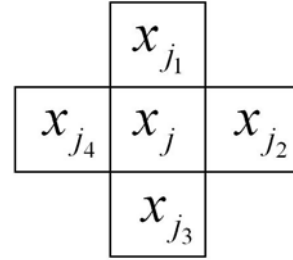


圖2 以一階鄰居集取得半監督式的樣本點

$$S_b = \alpha S_b^{LDA} + \beta S_b^{SLDA}$$

where define the between - spatial scatter matrix

$$S_b^{SLDA} = \sum_{i=1}^L \frac{1}{N_b(N_1 + \dots + N_L)} (m_i^s - m^s)(m_i^s - m^s)^T \quad (4)$$

$$S_w = \alpha S_w^{LDA} + \beta S_w^{SLDA}$$

where define the within - spatial scatter matrix

$$S_w^{SLDA} = \sum_{i=1}^L \frac{1}{N_b(N_1 + \dots + N_L)} \sum_{j=1}^{N_j} \sum_{\ell=1}^{N_b} \sum_{k=1}^{N_b} \sum_{g(x_{ik}^{(j)})=i} (x_{ik}^{(j)} - m_i^s)(x_{ik}^{(j)} - m_i^s)^T \quad (5)$$

其中，

$$\alpha, \beta \geq 0$$

$$\alpha, \beta \in \{0.5, 1, 1.5, 2\}$$

融入空間資訊最佳化下列準則可得到最佳的特徵(optimal features)：

$$J_{SLDA} = \text{trace}((S_w^{SLDA})^{-1} S_b^{SLDA}) \quad (6)$$

亦即求解廣義特徵值問題

$$S_b v_s = \lambda_s S_w v_s, \quad s=1, \dots, d \quad \text{with } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

3.3 融入空間資訊之演算法

本研究是以LDA為基礎，提出結合光譜與空間資訊之線性區別分析(SLDA)，在此將詳細介紹其演算法及架構示意圖。

$$N_i^s = \left\{ g(x_{\ell k}^{(j)}) = i \mid j = 1, 2, \dots, L; \right. \\ \left. \ell = 1, 2, \dots, N_j; k = 1, 2, \dots, N_b \right\}$$

N_i^s the set of samples in class i with spatial information

Let g be the decision function

N_b the samples of neighborhood

$$m_i^s = \frac{1}{N_i^s} \sum_{j=1}^L \sum_{\ell=1}^{N_j} \sum_{k=1}^{N_b} x_{\ell k}^{(j)}$$

the mean of the i with spatial information

$$m^s = \frac{1}{N_b N} \sum_{j=1}^L \sum_{\ell=1}^{N_j} \sum_{k=1}^{N_b} x_{\ell k}^{(j)}$$

the total mean with spatial information

SLDA 演算法描述如下：

Step1: 計算 N_i^s , training samples 的 neighborhood 個數

Step2: 使用 Fisher criterion

Step3: 計算 S_b^{SLDA} 與 S_w^{SLDA} · 得 Step2 最佳化的解

Step4: 解特徵值問題

$$S_b v_s = \lambda_s S_w v_s, \quad s = 1, \dots, d \quad \text{with } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$$

Step5: 從原始空間到降維子空間的轉移函數

$$A = [v_1, v_2, \dots, v_p] \quad p \leq d$$

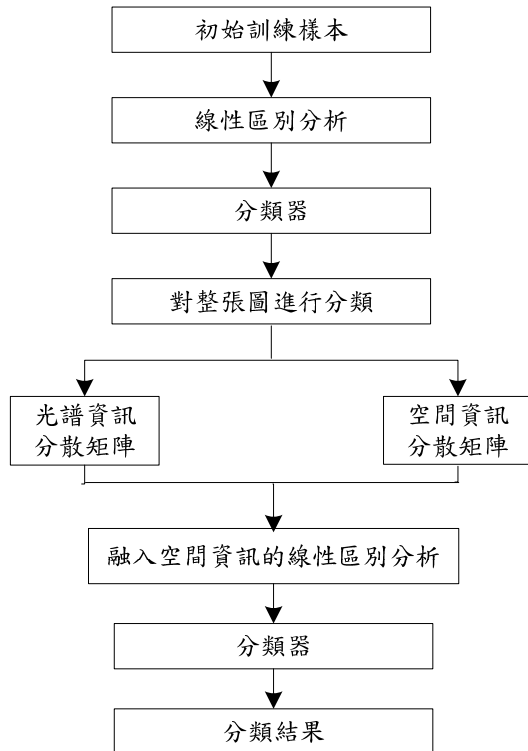


圖3 本研究架構之示意圖

4. 實驗設計

本實驗針對小樣本、高維度的情形，採用特徵萃取技術來解決 Hughes 現象，嘗試將空間資訊融入 LDA 的方法中，依據鄰居系統 (Neighbor System) 的概念，採用一階鄰居集取得半監督式的訓練樣本，經由迭代來更新樣本的類別，其預設的閾值為 0.001，最大的迭代數為 100，直至類別穩定收斂為止，並以參數 0.5、1、1.5、2 的組合來調整光譜及空間資訊的重要性。

實驗影像以每個類別隨機選取 20、40 和 300 個為初始訓練樣本及每類的測試樣本數皆為 100 個，以此來搭配傳統高斯分類器 (GC)、K 最近鄰演算法 (KNN)、支援向量機 (SVM) 之辨識器來進行此分類方式的效能驗證。

在分類器的部份，傳統高斯分類器 (GC)、K 最近鄰演算法 (KNN, K=1)、支援向量機 (SVM) 則以 One-against-all 為主，使用 RBF kernel 且參數 σ 使用網格法 (grid method) 的方式從 $\sqrt{0.005}$ 至 $\sqrt{5}$ 的範圍間隔 $\sqrt{0.05}$ 選取；而參數 C 設定範圍為

$$C \in \{0.1, 1, 10, 20, 60, 100, 160, 200, 1000\}.$$

4.1 實驗資料

本研究擬採用遙測領域相關期刊及研討會論文集內高光譜遙測影像辨識研究中常採用之高光譜遙測影像，此影像大多是可以免費獲得，第一個為都市區域影像「Washington DC Mall 資料集」，此影像具有 220 個頻譜，但部份頻譜被水吸收必須排除，因此僅具有 191 個有效頻譜，包含 7 個類別；第二個為農業用地影像「Indian Pine Site (IPS) 資料集」，此影像具有 220 個有效頻譜，包含 9 個類別。

本實驗將使用這兩張遙測高光譜影像資料進行方法的比較，表 1 為資料型態的描述。

表 1 遙測光譜影像資料

光譜影像(維度數) Washington DC Mall (191)

類別數	類別名稱	圖素數
1	Buildings	3834
2	Roads	680
3	Paths	616
4	Lawn	1928
5	Trees	919
6	Water	1224
7	Shadows	221
全部		9422



圖 4 Washington DC Mall 影像

光譜影像(維度數) Indian Pine Site (220)

類別數	類別名稱	圖素數
1	■ Corn – no till	1428
2	■ Corn - min till	830
3	■ Grass / Pasture	483
4	■ Grass / Trees	730
5	■ Hay – windrowed	478
6	■ Soybean - no till	972
7	■ Soybean – min till	2455
8	■ Soybean – clean till	593
9	■ Woods	1265
全部		9234



圖 5 Indian Pine Site 的 ground truth 影像

4.2 實驗結果

本研究為了評估訓練樣本的大小對維度之影響，所以分別使用三個 case 來分析探究： $N_i = 20 < N < d$ (case 1)、 $N_i = 40 < d < N$ (case 2) 及 $d < N_i = 300 < N$ (case 3)，每類的測試樣本數皆為 100 個。

表 2 The accuracy with two different datasets

($N_i = 20$, case 1)

Classifier	FE	Datasets Accuracy % (dimensionality)	
		IPS	DC
GC (QDC)	LDA	55.89% (7)	47.71% (3)
	SLDA	68.67% (8)	89.71% (5)
KNN (1NN)	LDA	60.22% (8)	64.43% (6)
	SLDA	71.89% (8)	90.29% (6)
SVM (OAA)	LDA	60.78% (8)	63.14% (6)
	SLDA	71.22% (8)	89.86% (5)

表 3 The accuracy with two different datasets

($N_i = 40$, case 2)

Classifier	FE	Datasets Accuracy % (dimensionality)	
		IPS	DC
GC (QDC)	LDA	69.00% (8)	86.29% (6)
	SLDA	77.78% (8)	91.29% (6)
KNN (1NN)	LDA	69.00% (8)	89.29% (6)
	SLDA	79.67% (8)	91.29% (6)
SVM (OAA)	LDA	69.44% (8)	89.00% (6)
	SLDA	81.11% (8)	91.14% (6)

表 4 The accuracy with two different datasets

($N_i = 300$, case 3)

Classifier	FE	Datasets Accuracy % (dimensionality)	
		IPS	DC
GC (QDC)	LDA	87.78% (8)	96.14% (5)
	SLDA	88.56% (8)	95.43% (6)
KNN (1NN)	LDA	86.00% (8)	95.86% (6)
	SLDA	86.22% (8)	96.00% (6)
SVM (OAA)	LDA	88.89% (8)	96.29% (6)
	SLDA	90.00% (8)	96.29% (6)

由表 2-4 的結果可知：在 Indian Pine Site 的資料集，最高的正確率分別為 71.89% (case1)、81.11% (case2)、90.00% (case3)；在 Washington DC Mall 的資料集，最高的正確率分別為 90.29% (case1)、91.29% (case2)、96.29% (case3)。

觀察表 2-4 可以清楚的發現，本研究所提的方法，無論應用在何種分類器或資料集上，其加入空間資訊皆可提高正確率，而其中又以 Washington DC Mall 的資料集在小樣本下的效果更佳。

由下列圖 6-7 可以看出本研究所提出的方法「結合光譜與空間資訊之線性區別分析(SLDA)」與「傳統的線性區別分析(LDA)」在搭配不同分類器的情況下，針對兩張影像三個不同的 case，所顯示辨識率最高的部份。

本研究藉由表格及圖像來詳細的呈現出實驗的結果，從中可以發現在有限的小樣本下，SLDA 皆可以得到較有效的分類成果。

5. 結論及未來研究方向

由上述實驗結果可知，在光譜資訊中融入空間資訊，的確能有效的改善分類的成果，故如何將空間資訊適當地融入特徵萃取的技術中，是個值得分析及探討的課題。

本實驗中，用以調整光譜及空間資訊的參數是以固定的組合方式實作，之後擬採用線性組合的方式，探討如何自動的挑選出最適合的參數，藉以提升分類的效能。

未來的研究方向為考慮將空間資訊融入其他的特徵萃取技術中，如 Nonparametric Weighted Feature Extraction (NWFE)或將空間資訊中採用鄰居系統(Neighbor System)的部份，改以 Graph 的概念來實作，或利用事前或事後機率來發展結合空間資訊之特徵萃取演算法。

參考文獻

- [1] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training". *In Proceedings of the Workshop on Computational Learning Theory*, Madison, WI, pp.92-100, 1998.
- [2] B-C. Kuo and D. A. Landgrebe, "A Covariance Estimator for Small Sample Size Classification Problems and Its Application to Feature Extraction," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 40, No. 4, pp 814-819, 2002.
- [3] B-C. Kuo and D. A. Landgrebe, "Regularized Covariance Estimators for Hyperspectral Data Classification and Its Application to Feature Extraction," *International Geoscience and Remote Sensing Symposium*, June 24-28, 2002.
- [4] Fukunaga, K. "Introduction to Statistical Pattern Recognition." San Diego, CA:Academic Press,1990.
- [5] Ghahramani, Z. and Jordan, M.I., "Learning from incomplete data". *Technical Report 108, MIT Center for Biological and Computational Learning*, 1994.
- [6] Hughes, G. F., "On the mean accuracy of statistical pattern recognition," *IEEE Trans. Inform. Theory*, vol. 14, pp. 55 - 63, Jan. 1968.
- [7] J. Crook and J. Banasik, "Sample selection bias in credit scoring models". *International Conference on Credit Risk Modeling and Decisioning*, Philadelphia, PA, 2002.
- [8] Jackson, Q. and Landgrebe, D.A., "An Adaptive Classifier Design for High-Dimensional Data Aanlysis with a Limited Training Data Set", *IEEE Geoscience and Remote Sensing*, Vol.39, No.

- 12, pp. 2664-2679, 2001.
- [9] J. Besag, “*On the statistical analysis of dirty pictures*, *J. Royal Statist.*” Soc., vol. 68, pp.259-302, 1986.
- [10] J. Besag, “*Spatial interaction and the statistical analysis of lattice systems.*” *J. Royal Statist. Soc.*, vol. 36, no. 2, pp. 192-236, 1974.
- [11] Miller, D. and Uyar, S., “A mixture of experts classifier with learning based on both labeled and unlabeled data” *data in Adv. in Neural Information Processing Systems 9*.
- [12] R. Kinderman and J. L. Snell, “*Markov random fields and their applications*,” Amer. Math. Soc., vol. 1, pp. 1-142, 1980.
- [13] S. Goldman and Y. Zhou, “Enhancing supervised learning with unlabeled data”. *In Proceedings of the Seventeenth International Conference on Machine Learning*, San Francisco, CA, pp.327-334, 2000.
- [14] S. German and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration,” *IEEE Trans. Pat. An. Mach, Intell*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.
- [15] S. Z. Li, “*Markov Random Field Modeling in Computer Vision*,” Berlin, Germany: Spring-Verlag, 1995.

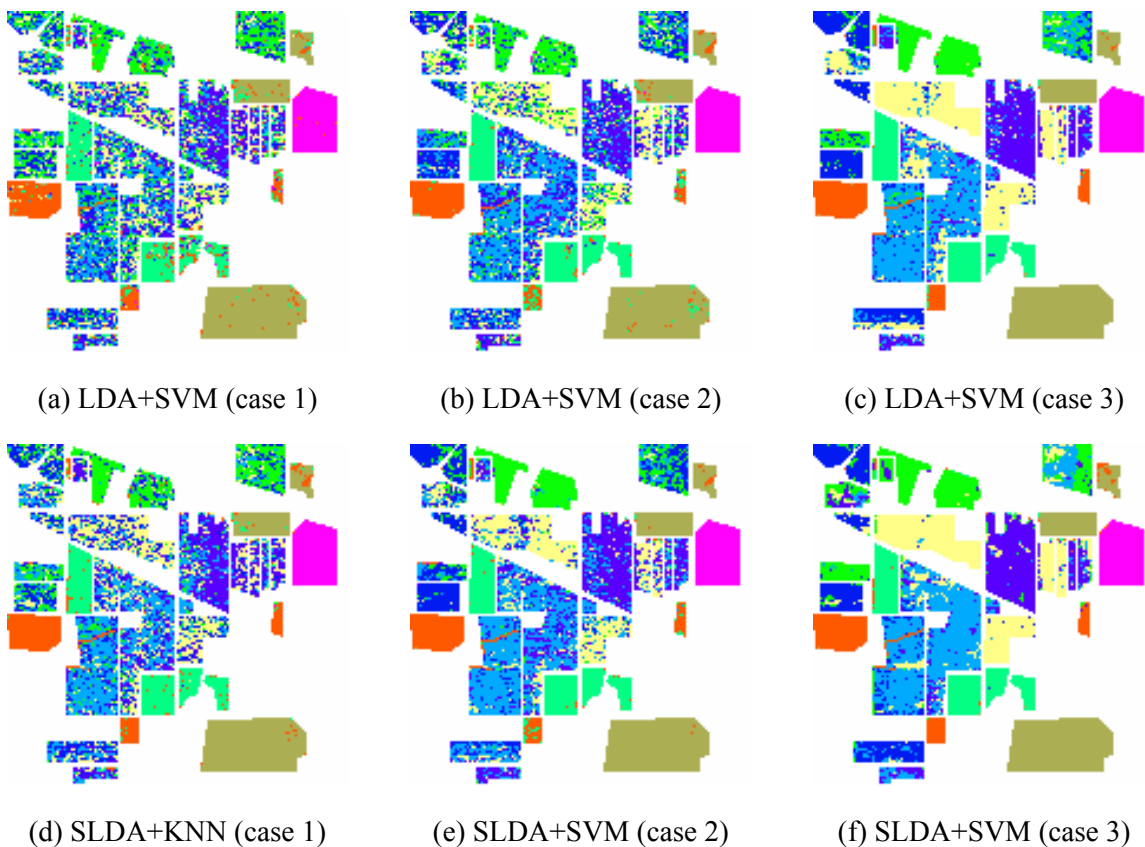
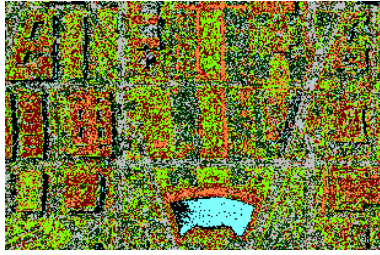
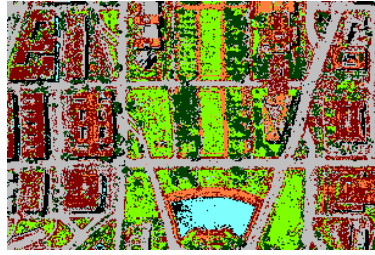


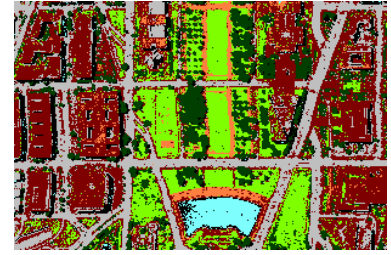
圖6 圖(a)-(c)為在Indian Pine Site dataset上，使用LDA搭配SVM的thematic maps。(d)-(f)為在Indian Pine Site dataset上，使用SLDA搭配KNN或SVM的thematic maps。



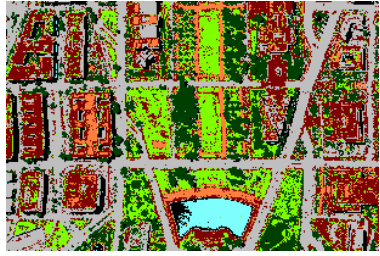
(g) LDA+KNN (case 1)



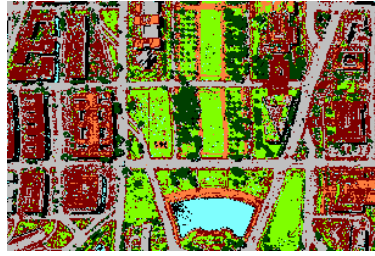
(h) LDA+KNN (case 2)



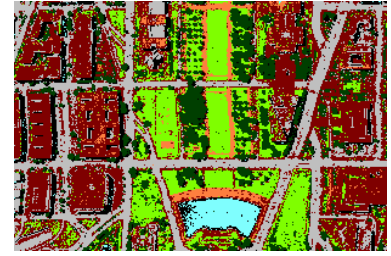
(i) LDA+SVM (case 3)



(j) SLDA+KNN (case 1)



(k) SLDA+KNN (case 2)



(l) SLDA+SVM (case 3)

圖7 (g)-(i)為在Washington DC Mall dataset上，使用LDA搭配KNN或SVM的thematic maps。
(j)-(l)為在Washington DC Mall dataset上，使用SLDA搭配KNN或SVM的thematic maps。