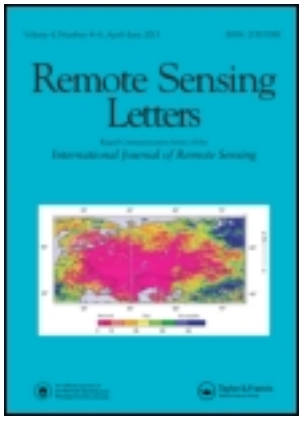

Remote Sensing Letters
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/trsl20

Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery

Bei Zhao[a], Yanfei Zhong[a] & Liangpei Zhang[a]
[a] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, 129 Luoru Road, Wuhan, P.R. China
Published online: 20 Nov 2013.

To cite this article: Bei Zhao, Yanfei Zhong & Liangpei Zhang (2013) Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery, Remote Sensing Letters, 4:12, 1204-1213

To link to this article: http://dx.doi.org/10.1080/2150704X.2013.858843

Remote Sensing Letters, 2013
Vol. 4, No. 12, 1204–1213, http://dx.doi.org/10.1080/2150704X.2013.858843

Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery

BEI ZHAO, YANFEI ZHONG* and LIANGPEI ZHANG
State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, 129 Luoru Road, Wuhan, P.R. China

(*Received 29 July 2013; accepted 17 October 2013*)

In order to capture the high-level concepts in high spatial resolution (HSR) remote sensing imagery, scene classification based on a latent Dirichlet allocation (LDA) model, a generative topic model, is a practical method to bridge the semantic gaps between the low-level features and the high-level concepts of HSR imagery. In the previous work, LDA has been considered as a scene classifier, namely C-LDA, and multiple LDA models for each scene class are built separately, where the scene class is determined by a maximum likelihood rule. The C-LDA strategy disregards the correlations between the generative topic spaces of the different scene classes. In this letter, two novel strategies of scene classification based on LDA are proposed to consider the correlations between the generative topic spaces of the different scene classes by sharing the topic spaces for all the scene classes. One of the proposed strategies utilizes LDA as part of the classifier, namely P-LDA, which generates the topic space from all the training images. A discriminative classifier (e.g., support vector machine, SVM) is also employed as the other classification part of P-LDA. The other proposed strategy employs LDA as the topic feature extractor, namely F-LDA, which generates the topic space from all the training and test images, and utilizes a discriminative classifier to classify the topic features. The experimental results using aerial orthophotographs show that the performances of the two proposed strategies for scene classification based on LDA are better than the traditional C-LDA method.

1. Introduction

In recent years, a large amount of high spatial resolution (HSR) remote sensing images with abundant spatial details has become available, allowing precise earth land-use/land-cover investigation (Batista and Haertel 2010, Kim *et al.* 2011). Object-based image analysis and contextual-based classification methods have been used to extract and recognize the land-cover objects, such as buildings, trees, grass, and roads, for HSR remote sensing imagery (Bruzzone and Carlin 2006, Blaschke 2010, Pu *et al.* 2011, Tilton *et al.* 2012, Lizarazo 2013). However, there are often semantic gaps between the objects (e.g., buildings) and the high-level semantic concepts in the images (e.g., residential areas or industrial areas). To solve this problem, semantic scene classification methods, such as the Bayesian framework

*Corresponding author. Email: zhongyanfei@whu.edu.cn

for object recognition (Aksoy *et al.* 2005), have been proposed for remote sensing imagery. Unlike the Bayesian method for scene modelling (Aksoy *et al.* 2005), and the discriminative model with bag-of-words feature (Zhou *et al.* 2013), the family of latent generative topic models, such as probabilistic latent semantic analysis (pLSA) (Hofmann 2001) and latent Dirichlet allocation (LDA) (Blei *et al.* 2003), has been successfully utilized to model scenes without object recognition for natural scene imagery (Li and Perona 2005, Bosch *et al.* 2008, Zhou *et al.* 2013). Compared to the classification methods directly using the bag-of-words, the latent generative topic model, pLSA, performs better due to the latent topic space being generated by the generative model (Bosch *et al.* 2008). Nevertheless, pLSA tends to overfit the samples as the number of training samples increases (Zhou *et al.* 2013). LDA overcomes this weakness and is a well-defined generative model. Recently, some refined models, such as correlated LDA, supervised LDA, and maximum entropy discriminative LDA, have been proposed to cope with textual analysis or natural image analysis, based on the LDA model (Blei and Lafferty 2007, Blei and McAuliffe 2007, Zhu *et al.* 2012). To cope with HSR image scene classification, LDA has been utilized to bridge the semantic gaps by building multiple LDA models for each scene class and determining the scene class using the maximum likelihood rule (Liénou *et al.* 2010). In the previous work (Liénou *et al.* 2010), the LDA model is regarded as the scene classifier, namely C-LDA. Although C-LDA obtains satisfactory scene classification results, it misses the important correlations between the latent topic spaces of the different scene classes, because of the separately built models of the multiple scenes.

To improve the scene classification ability, two hybrid generative/discriminative scene classification strategies are proposed to sufficiently utilize the correlations between the latent topic spaces of the different scene classes. In the first strategy, LDA as a part of the scene classifier, namely P-LDA, is designed to utilize the LDA trained by the training images to classify the test images into the topic space generated from all the scene classes, and uses a discriminative model (such as support vector machine, SVM) as the other part of the scene classifier. The other strategy, with LDA as a feature extractor, namely F-LDA, is utilized to extract the topic features from all the training and test images, and employs a discriminative model to classify the features. In both the P-LDA and F-LDA strategies, all the scene classes share the same latent topic space. In contrast to P-LDA, F-LDA generates the latent topic space using all the training and test images, instead of just the training images. By combining the LDA generative model with the other discriminative model, the two proposed classification strategies can utilize the correlations between the topic spaces of the different scene classes. The experimental results show that P-LDA and F-LDA can obtain better scene classification accuracy than C-LDA.

2. Previous related work

In this section, the generation process, inference, and parameter estimation of LDA (Blei *et al.* 2003) are briefly described. The strategy of scene classification utilizing LDA as the scene classifier (Liénou *et al.* 2010), namely C-LDA, is also introduced.

2.1 *The LDA model*

In image processing, the basic processing element of LDA is the "word", which can be a pixel, a window of pixels (tile), or a segment (region) of the image. Given the

image set with M images $D = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_M\}$, the features of the processing elements in the images will be extracted and vector quantized into V clusters before LDA processing. Each image can then be described as a sequence of N cluster labels, denoted by $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)$.

For the LDA generation process, the topic $z_{d,n}$ of the processing elements $w_{d,n}$ in the image $\boldsymbol{w}_d$ are generated from a multinomial distribution with parameters $\boldsymbol{\theta}_d$. $\boldsymbol{\theta}_d$ follows a Dirichlet distribution with the prior $\boldsymbol{\alpha}$. The probability of $w_{d,n}$ being conditioned on the topic $z_{d,n}$ is $P(w_{d,n} \mid z_{d,n}, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a matrix recording the probabilities of each topic generating each processing element. The likelihood of the dth image $\boldsymbol{w}_d$ is written in equation (1), where n is the index of the processing elements in an image, N_d is the number of the processing elements in the dth image.

$$P(\boldsymbol{w}_d|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int P(\boldsymbol{\theta}_d|\boldsymbol{\alpha}) \left(\prod_{n=1}^{N_d} \sum_{z_{d,n}} P(z_{d,n}|\boldsymbol{\theta}_d) P(w_{d,n}|z_{d,n}, \boldsymbol{\beta}) \right) \mathrm{d}\boldsymbol{\theta}_d \tag{1}$$

To maximize the likelihood of the entire image set D, approximate variational inference technology is employed to estimate and infer the LDA model (Blei *et al.* 2003).

During the parameter estimation of LDA, the model parameter $\boldsymbol{\beta}$ can be obtained, and parameter $\boldsymbol{\alpha}$ can be fixed or updated using the Newton–Raphson method (Blei *et al.* 2003). In this letter, parameter $\boldsymbol{\alpha}$ is fixed. For the inference of LDA, the approximate probability of the image $P(\boldsymbol{w}_d \mid \boldsymbol{\alpha}, \boldsymbol{\beta})$ can be acquired using $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

2.2 *The C-LDA strategy for scene classification*

To achieve the scene classification, C-LDA (Liénou *et al.* 2010) views the LDA model as the scene classifier. The procedure of C-LDA (figure 1) is described as follows.

First, the features of the basic processing elements in the images are extracted and quantized into V clusters by the k-means algorithm. Second, L LDA models with

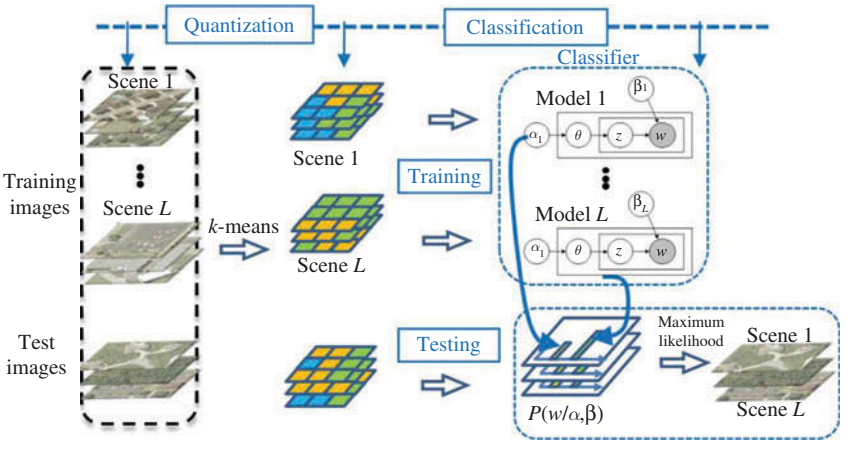

Figure 1. C-LDA strategy of scene classification based on LDA.

$\{(\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s) \mid s = 1, \ldots, L\}$ are trained for the L scene classes. Finally, the approximate probability of the test image $P(\boldsymbol{w} \mid \boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$ is inferred using each scene class LDA model parameter $(\boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$. The scene class label of $\boldsymbol{w}$ is determined by the maximum likelihood rule: $s^* = \arg\max_s P(\boldsymbol{w} \mid \boldsymbol{\alpha}_s, \boldsymbol{\beta}_s)$.

From the procedure of the C-LDA strategy, it can be seen that the LDA models for the scene classes are built separately, and each scene class owns a topic space generated by the corresponding LDA model. However, this strategy disregards the correlations between the topic spaces of the different scene classes.

3. Hybrid generative/discriminative scene classification strategies

To sufficiently consider the correlations between the topic spaces of the different scene classes, two hybrid generative/discriminative classification strategies, P-LDA and F-LDA, are proposed. Instead of using the approximate probability $P(\boldsymbol{w} \mid \boldsymbol{\alpha}, \boldsymbol{\beta})$ of the image for the scene classification, both P-LDA and F-LDA employ the variational parameter $\boldsymbol{\gamma}$ in the variational inference of LDA (Blei *et al.* 2003) as the topic features to describe the images. During the variational inference, the updating of the variational parameter $\boldsymbol{\gamma}$ is shown in equation (2), while the updating of the other variational parameters $\Phi = \{\phi_{d,n,i} \mid d = 1, \ldots, M; n = 1, \ldots, N_d; i = 1, \ldots, K\}$ is shown in equation (3). K is the number of the topics.

$$\gamma_{d,i} = \alpha_i + \sum_{n=1}^{N_d} \phi_{d,n,i} \tag{2}$$

$$\phi_{d,n,i} \propto \boldsymbol{\beta}_{i,w_n} \left(\psi(\gamma_{d,i}) - \psi\left(\sum_{j=1}^{K} \gamma_{d,j} \right) \right) \tag{3}$$

In equation (2), $\phi_{d,n,i}$ is a variational multinomial parameter recording the probability of $w_{d,n}$ in $\boldsymbol{w}_d$ belonging to the ith topic, ψ is the first derivative of the log Gamma function, and $\boldsymbol{\beta}_{i,w_n}$ is a component of model parameter $\boldsymbol{\beta}$ and is equivalent to $P(w_{d,n} \mid z_{d,n} = i)$. α_i in equation (3) is a component of model parameter $\boldsymbol{\alpha}$, and $\gamma_{d,i}$ is a variational Dirichlet parameter with respect to the ith component of the topic distribution for the image $\boldsymbol{w}_d$. It should be noted that the parameter $\gamma_{d,i}$ is image-specific and can be viewed as a representation of the image in the topic space. After the inference of LDA, the topic features $\boldsymbol{\gamma}_d = (\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,K})$ can be acquired. The procedures of the application of the topic features $\boldsymbol{\gamma}_d$ in P-LDA and F-LDA are described in the following two parts.

3.1 *The P-LDA strategy*

In the P-LDA strategy (figure 2), the generative LDA model and the discriminative SVM model are combined to form a cascaded classifier. The LDA model supplies the SVM model with topic features, and the final scene class label is obtained by SVM. The procedure of P-LDA is illustrated in figure 2.

1. **Quantization:** This step is the same as the quantization step in C-LDA, where the feature extracted from the basic elements is clustered into V bins by the

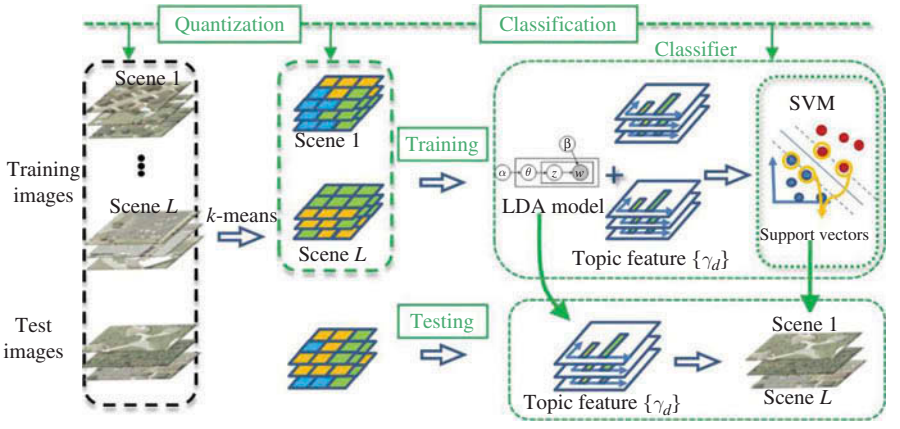

Figure 2. P-LDA strategy of scene classification based on LDA.

k-means method. All the training images and test images are transformed into the cluster labels (figure 2).

2. **Training:** All the training images are used to train one LDA model. In the training process, the LDA model with ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$), and the topic features $\{\boldsymbol{\gamma}_d\}$ of all the training images, are obtained by equations (2) and (3). Meanwhile, the topic features $\{\boldsymbol{\gamma}_d\}$ are used to train the SVM model. With the coordination of the LDA model and the SVM model, the corresponding relationships between the observed image $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)$, the topic features, and the scene class labels are built. After the training procedures of the LDA model and SVM model are accomplished, the training step of the hybrid classifier is complete.
3. **Testing:** For a test image $\boldsymbol{w}_d = (w_1, w_2, \ldots, w_N)$, the topic features $\boldsymbol{\gamma}_d$ will be acquired using the LDA model of the trained hybrid cascaded classifier. Subsequently, the topic features $\boldsymbol{\gamma}$ will be classified by the other part of the cascaded classifier, SVM, to determine the scene class.

By the application of the topic features $\{\boldsymbol{\gamma}_d\}$, P-LDA has the capability to share the topic space for all the scene classes.

3.2 *The F-LDA strategy*

Compared to forming the cascaded classifier using LDA and SVM in P-LDA, the proposed F-LDA (figure 3) views LDA as the feature extractor and extracts the topic features for all the images.

1. **Quantization:** This step is the same as (1) in P-LDA. By use of the *k*-means method, all the images are quantized into cluster labels, which are used for the further processing.
2. **Topic feature extraction:** In this step, the cluster label sequences of all the images in D are used to generate the topic features $\{\boldsymbol{\gamma}_d\}$ by the estimation procedure of the LDA model.
3. **Classification:** The topic features of all the training images are used to train the SVM classifier. The scene class of the test image $\boldsymbol{w}$ is determined by

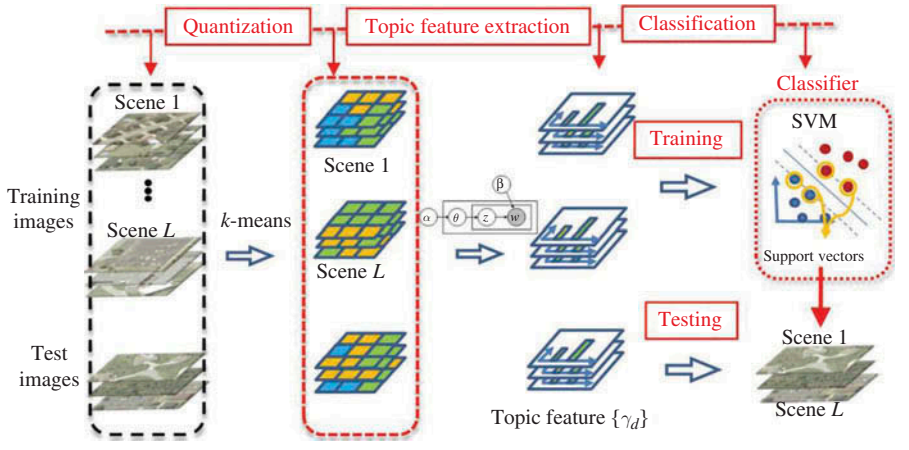

Figure 3. F-LDA strategy of scene classification based on LDA.

classifying the topic features of the test image $\boldsymbol{w}$ using the trained SVM classifier.

By considering the LDA model as a feature extractor, F-LDA can generate a shared topic space for all the images.

4. Experimental results and analysis

4.1 *Data sets and parameter settings*

In this section, a series of experiments is designed to compare the performance of C-LDA, P-LDA, and F-LDA. The experimental data set is constructed from 100 aerial orthophotographs with a spatial resolution of 0.61 m, acquired from the USGS, covering Montgomery County, Ohio, USA. This data set contains 143, 133, 100, and 139 sample images with a size of 150 pixels × 150 pixels for four scene classes: residential area (RA), farm (FA), forest (FO), and car park (CP), respectively (figure 4).

In the experiments, the features extracted from the basic elements consist of the means and standard variations of the basic elements, which will be quantized using the *k*-means clustering algorithm. For the hybrid generative/discriminative strategies, P-LDA and F-LDA, the SVM model is selected as the discriminative model, and the

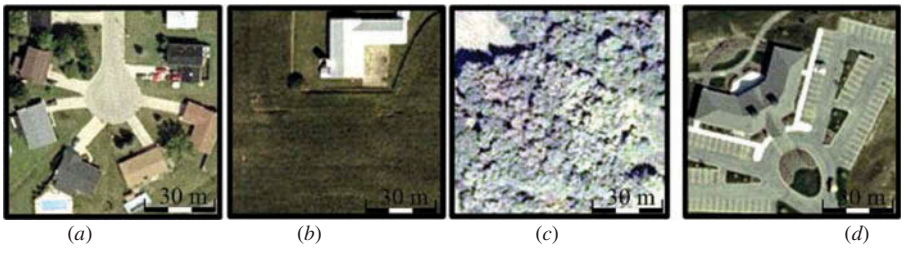

Figure 4. Four scene classes of aerial orthophotographs shown with natural colour, obtained from the USGS, covering Montgomery County, Ohio, USA. (*a*) RA. (*b*) FA. (*c*) FO. (*d*) CP.

training images are randomly selected from the whole image set. The topic number K is varied from 10 to 70 with a step size of 5, and is selected according to equation (4) by 3-fold cross validation technology. It is worth noting that the perplexity $F(D)$ in equation (4) decreases as the likelihood $P(\boldsymbol{w}_d)$ increases, which means that a higher value of $F(D)$ is better. The classification results are evaluated by the confusion matrix and overall accuracy. For each set of parameters, the scene classification is executed 10 times to obtain the mean and standard deviation of the classification accuracy.

$$F(D) = \exp\left(-\sum_{d=1}^{M} \ln P(\boldsymbol{w}_d) / \sum_{d=1}^{M} N_d\right) \tag{4}$$

4.2 *Experimental results*

With 50 images and a window size of 8 pixels × 8 pixels, the best classification accuracies of C-LDA, P-LDA, and F-LDA are reached when the corresponding number of clusters V is equal to 100, 100, and 400, respectively. The classification accuracies are reported in table 1. From table 1, it can be seen that P-LDA improves the accuracy by about 3%. The corresponding confusion matrixes of C-LDA, P-LDA, and F-LDA are shown in tables 2(*a*), (*b*), and (*c*), respectively. The experimental results shown in tables 2(*a*), (*b*), and (*c*), infer that P-LDA and F-LDA improve the accuracy mainly in the FA scene class.

A set of experiments, with the number of clusters V varied from 100 to 500, are conducted to test the influence of the cluster number when the number of training images and window size are set to 50 and 8 pixels × 8 pixels, respectively (figure 5(*a*)). Figure 5(*a*) shows that the proposed P-LDA and F-LDA outperform the C-LDA as the number of clusters varies from 100 to 500, while the accuracy of F-LDA is a little higher than P-LDA. The corresponding computation times are reported in figure 5(*b*), which indicates that the computation cost of F-LDA is almost three times as much as P-LDA and C-LDA. The computation costs of P-LDA and C-LDA are comparable.

In order to evaluate the effect of the window size, scene classifications with three different window sizes (5 pixels × 5 pixels, 8 pixels × 8 pixels, and 10 pixels × 10 pixels) are performed using C-LDA, P-LDA, and F-LDA. The number of clusters V and the number of training samples are set to 200 and 50, respectively. The results are presented in figure 5(*c*), which indicates that P-LDA and F-LDA perform better than C-LDA with window sizes of 8 pixels × 8 pixels and 10 pixels × 10 pixels. Although C-LDA performs better than P-LDA and F-LDA with a window size of 5 pixels × 5 pixels, the classification accuracy is lower than the accuracy of P-LDA and F-LDA with window sizes of 8 pixels × 8 pixels and 10 pixels × 10 pixels.

Table 1. The best classification accuracies of C-LDA, P-LDA, and F-LDA with 50 training images and a window size of 8 pixels × 8 pixels.

Method	C-LDA	P-LDA	F-LDA
Accuracy	0.925 ± 0.023	0.957 ± 0.0178	0.959 ± 0.017

Table 2. Confusion matrixes of the three strategies. Numbers in the tables represent the numbers of images. (*a*) C-LDA (overall accuracy 93.3%). (*b*) P-LDA (overall accuracy 95.9%). (*c*) F-LDA (overall accuracy 96.2%).

(*a*)

		Classification				
Classes		RA	FA	FO	CP	Total
Reference data	RA	92	10	0	4	106
	FA	0	67	0	0	67
	FO	0	1	50	0	51
	CP	1	5	0	85	91
	Total	93	83	50	89	315

(*b*)

		Classification				
Classes		RA	FA	FO	CP	Total
Reference data	RA	89	1	0	7	90
	FA	1	81	0	0	91
	FO	0	0	50	0	50
	CP	3	1	0	82	84
	Total	93	83	50	89	315

(*c*)

		Classification				
Classes		RA	FA	FO	CP	Total
Reference data	RA	86	0	0	4	97
	FA	7	83	0	1	82
	FO	0	0	50	0	50
	CP	0	0	0	84	86
	Total	93	83	50	89	315

A further set of experiments are designed to analyse the impact of the number of training samples, where the training images for each scene class are randomly selected (figure 5(*d*)). From figure 5(*d*), it can be seen that the classification accuracies of P-LDA and F-LDA are higher than C-LDA, while the accuracy of F-LDA is the highest.

5. Conclusions

In this letter, two hybrid generative/discriminative scene classification strategies for scene classification based on LDA are proposed to consider the correlations between the generative topic spaces of the different scene classes by sharing the topic spaces for all the scene classes. One of the proposed strategies, P-LDA, utilizes generative model LDA as part of the classifier to generate the topic space over all the training images, and employs another discriminative classifier (e.g., SVM) as the other part of

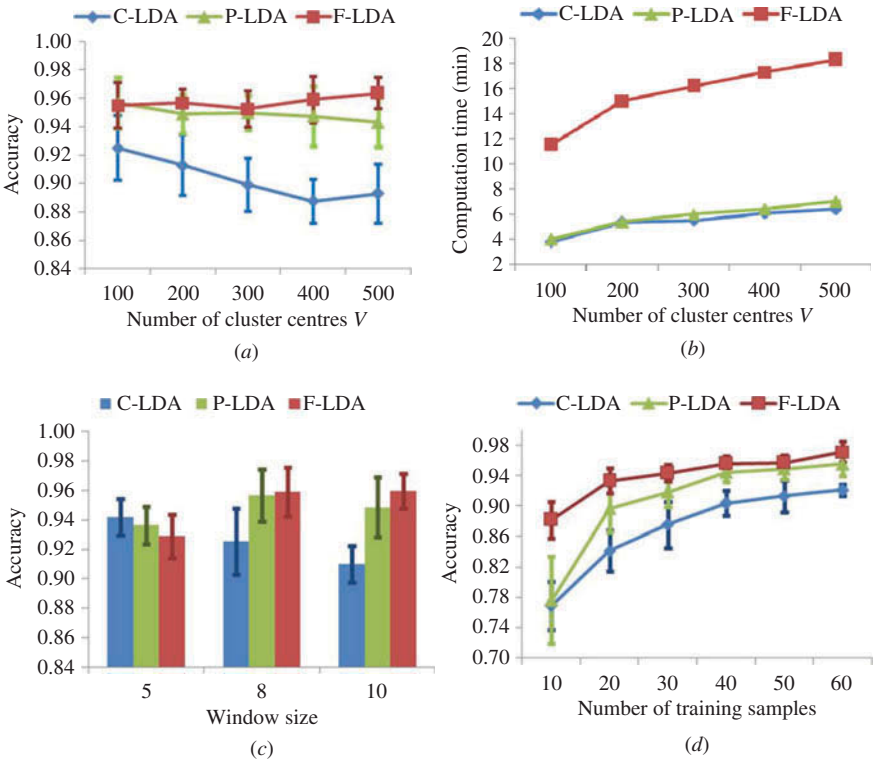

Figure 5. Sensitivity analysis for the performance of the three strategies. (*a*) Scene classification with different numbers of clusters. (*b*) Computation times with different numbers of clusters. (*c*) Scene classification with different window sizes. (*d*) Scene classification with different numbers of training samples.

P-LDA. The other proposed strategy, F-LDA, employs LDA as the topic feature extractor to generate the topic space from all the training and test images, and utilizes another discriminative classifier to classify the topic features. Compared to the C-LDA strategy, which views LDA as the scene classifier and builds multiple LDA models for each scene class separately, the experiments show that both proposed strategies for scene classification based on LDA perform better than C-LDA. Through the analysis of the effect of the cluster number on the scene classification, it can be seen that the proposed hybrid generative/discriminative strategies, P-LDA and F-LDA, tend to obtain better classification accuracies. Meanwhile, with the different numbers of clusters, the computation cost of F-LDA is greater than P-LDA and C-LDA, and the computation times of P-LDA and C-LDA are comparable. By analysing the effect of the window size and the number of training samples, the proposed strategies, P-LDA and F-LDA, both show stable characteristics.

Acknowledgements

The authors would like to thank the editor, associate editor and anonymous reviewers for their helpful comments and suggestions.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 41371344], and Foundation for the Author of National Excellent Doctoral Dissertation of P.R. China (FANEDD) [grant number 201052].

References

Aksoy, S., Koperski, K., Tusk, C., Marchisio, G. and Tilton, J.C., 2005, Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, pp. 581–589.

Batista, M.H. and Haertel, V., 2010, On the classification of remote sensing high spatial resolution image data. *International Journal of Remote Sensing*, **31**, pp. 5533–5548.

Blaschke, T., 2010, Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, **65**, pp. 2–16.

Blei, D.M. and Lafferty, J.D., 2007, A correlated topic model of science. *The Annals of Applied Statistics*, **1**, pp. 17–35.

Blei, D.M. and McAuliffe, J.D., 2007, Supervised topic models. In *Twenty-First Annual Conference on Neural Information Processing Systems*, 3–6 December 2007, Vancouver, BC (Cambridge, MA: MIT Press), pp. 121–154.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003, Latent Dirichlet allocation. *Journal of Machine Learning Research*, **3**, pp. 993–1022.

Bosch, A., Zisserman, A. and Muñoz, X., 2008, Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, pp. 712–727.

Bruzzone, L. and Carlin, L., 2006, A multilevel context-based system for classification of very high spatial resolution Images. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2587–2600.

Hofmann, T., 2001, Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, **42**, pp. 177–196.

Kim, M., Warner, T.A., Madden, M. and Atkinson, D.S., 2011, Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing*, **32**, pp. 2825–2850.

Li, F.-F. and Perona, P., 2005, A Bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 20–26 June 2005, San Diego, CA (Washington, DC: IEEE Computer Society), pp. 524–531.

Liénou, M., Maître, H. and Datcu, M., 2010, Semantic annotation of satellite images using latent Dirichlet allocation. *IEEE Geoscience and Remote Sensing Letters*, **7**, pp. 28–32.

Lizarazo, I., 2013, Meaningful image objects for object-oriented image analysis. *Remote Sensing Letters*, **4**, pp. 419–426.

Pu, R., Landry, S. and Yu, Q., 2011, Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. *International Journal of Remote Sensing*, **32**, pp. 3285–3308.

Tilton, J.C., Tarabalka, Y., Montesano, P.M. and Gofman, E., 2012, Best merge region-growing segmentation with integrated nonadjacent region object aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, **50**, pp. 4454–4467.

Zhou, L., Zhou, Z. and Hu, D., 2013, Scene classification using a multi-resolution bag-of-features model. *Pattern Recognition*, **46**, pp. 424–433.

Zhu, J., Ahmed, A. and Xing, E.P., 2012, MedLDA: maximum margin supervised topic models. *Journal of Machine Learning Research*, **13**, pp. 2237–2278.