AN OBJECT-ORIENTED CLUSTERING ALGORITHM FOR MULTI-SPECTRAL IMAGES USING THE LATENT DIRICHLET ALLOCATION

Hong Tang [a, *], Yingfeng Qi [a], Li Shen [a], Yang Shu [a]

[a] ADREM, Beijing Normal University, 100875 Beijing, P. R. China

KEY WORDS: Latent Dirichlet allocation, Object-oriented clustering, Multi-spectral images

ABSTRACT:

In this paper, we present a novel framework to integrate the object-oriented clustering process into the modelling of the Latent Dirichlet Allocation for multi-spectral remote sensing images. In the framework, a word (i.e., *pixel*) should be allocated a same topic label during the modelling no matter what document (i.e., *sub-image*) it resides in. It differs from state-of-the-art ways where a two-separate-step is utilized to cluster VHR satellite images: (1) modelling: the latent Dirichlet Allocation and its relative are applied to model a collection of overlapped sub-images (i.e., documents), which are partitioned (or segmented) from a satellite image; (2) clustering: a classification map is derived from the learned model by combining multiple labels of each pixel into a unique one. Experimental results over a multi-spectral image show that the proposed method could well separate two geo-objects whose grade values are very similar to each other, i.e., the *shallow* and *water*.

1. INTRODUCTION

The probabilistic topic models, as a set of language models [1-2], have been employed to classify [3-4], detect [5] and cluster [6] geo-objects in remote sensing images. Unlike modeling text using probabilistic topic models, one needs to design both words and documents from remote sensing images before modeling [3-7]. The words used to be derived by clustering visual features of a pixel or a set of pixel in a squared area. A segment or a sub-image with pre-fixed size is often used as a document. To overcome the limitation of the *bag-of-word* assumption, there exists overlapped area over adjacent documents, and a classification label a combination of all of documents who include the pixel.

In this paper, we present a novel framework to integrate the object-oriented clustering process into the modelling of the Latent Dirichlet Allocation (LDA) for multi-spectral remote sensing images. In the framework, a pixel should be allocated a same topic label during the modelling no matter what document (i.e., *sub-image*) it resides in. In addition, unlike the existed works, there exist multiple words on a pixel in a document. Specifically, a visual word is allocated to each pixel of every band in a multi-spectral image. For the sake of simplification, the proposed algorithm is called as a unified Latent Dirichlet Allocation (uLDA).

2. LATENT DIRICHLET ALLOCATION

2.1 Notation

Given a satellite image **I** with size of R rows and C columns, let $\mathcal{T}=\{ t_{ij}= (i\text{-}1)*C+j \mid 1\leqslant R, 1\leqslant R\leqslant C\}$ be the set of lattice sites in the image. A random field indexed by the lattice system $\mathcal{T}$ is given by $\mathbf{X} =\{ X_t= x_t \mid x_t \in \mathcal{A}, t \in \mathcal{T} \}$, where a random variable X_t at site t takes a value x_t in its state space $\mathcal{A}=\{1, 2, \cdots, d\}$. The set x = $\{x_t \mid t \in \mathcal{T}\}$ is a sample drawn from the state space $\mathcal{X}$ with the joint probability P(X = x). Given an observed sample x of a pixel attribute (e.g., spectra or texture) random field **X**, the goal of image clustering is to obtain a sample $\mathbf{c} = \{c_t \mid t \in \mathcal{T} \}$ of a label random field **C** according to a criterion, e.g. maximum a posterior P(**C** | **X**).

Unlike pixel-based clustering method, the LDA is to model a collection of documents (i.e., sub-images), which includes a set of words (e.g., attributes of pixels). Without loss of generality, we assume that each pixel is covered by a squared sub-image with size of $h \times h$. Therefore, a collection of documents indexed by $\mathcal{T}$ is given by $\mathcal{D}_0 = \{ d_t = \vec{w}_t \mid t \in \mathcal{T}\}$, where the document d_t centering at site t is a sub-image including pixels whose distance to the site t in either row or column is not larger than $h/2$. If gray values in a panchromatic image are regarded as visual words, t-th document could be given by $\vec{w}_t = \{x_{t_1}, \cdots, x_{t_N}\}$.

Given a collection of documents $\mathcal{D}_0$, the LDA approaches the clustering in a two-step way: (1) modeling: all of the sub-images are modeled by using the LDA. Consequently, every pixel in each sub-image would be allocated a topic label; (2) clustering: The cluster label of every pixel in image **I** is derived by combining multiple topic labels of the pixel, which scatters in multiple sub-images.

2.2 Modeling using the Latent Dirichlet Allocation

As shown in Fig. 1 (a), the LDA assumes the following generative process for i-th document $d_i = \{w_{i1}, w_{i2}, \cdots, w_{iN}\}$ in a corpus $\mathcal{D}_0 = \{ d_t \mid t \in \mathcal{T}\}$:

1) Sample $\theta_i \sim$ dir (α).
2) Sample $\Phi\sim$dir(β).
3) For each of the N words $\{w_{i1}, w_{i2}, \cdots, w_{iN}\}$.
   (a) Sample a topic $z_{ij} \sim$ *multinomial* (θ_i).
   (b) Sample a word $w_{ij} \sim P(w_{ij} \mid z_{ij}, \Phi)$ a multinomial probability conditioned on the topic z_{ij}.

* Corresponding author: tanghong@bnu.edu.cn; Tel.: +86-10-5880 6401.

The key inferential problem in the LDA is that of computing the posterior distribution of the hidden variables given a document

$$P(\vec{z}_i \mid \vec{w}_i, \alpha, \beta) = \frac{P(\vec{z}_i, \vec{w}_i \mid \alpha, \beta)}{P(\vec{w}_i \mid \alpha, \beta)} \quad (1)$$

where the joint distribution $P(\vec{z}_i, \vec{w}_i \mid \vec{\alpha}, \vec{\beta})$ could be approximated by interactively sampling each of the N topics

$$\begin{aligned} & P(z_{ij}, w_{ij} \mid \vec{z}_{i\neg j}, \vec{w}_{i\neg j} \vec{\alpha}, \vec{\beta}) \\ & = \frac{P(\vec{z}_{ij}, \vec{w}_{ij} \mid \vec{\alpha}, \vec{\beta})}{P(\vec{z}_{i\neg j}, \vec{w}_{i\neg j} \mid \vec{\alpha}, \vec{\beta})} \\ & = \frac{n_i^{(k)} + \alpha_k}{[\sum_{k=1}^{K}(n_i^{(k)} + \alpha_k)] - 1} * \frac{n_k^{(j)} + \beta_j}{[\sum_{j=1}^{V}(n_i^{(j)} + \beta_j)] - 1} \end{aligned} \quad (2)$$

Given a sample of $P(\vec{z}_i, \vec{w}_i \mid \alpha, \beta)$ for all of documents, the parameters of the empirical Bayes are given by

$$\theta_{ik} = \frac{n_i^{(k)} + \alpha_k}{\sum_{k=1}^{K}(n_i^{(k)} + \alpha_k)} \quad (3)$$

$$\phi_{kj} = \frac{n_k^{(j)} + \beta_j}{\sum_{j=1}^{V}(n_i^{(j)} + \beta_j)} \quad (4)$$

where V and K are the numbers of words in the vocabulary and topics, respectively. For the detail of the derivation, please refer to [2].

2.3 Clustering by combining multiple topic labels

Since documents are assumed to be independent in the LDA, a same pixel in the image **I** may be allocated different topic labels when it resides in different documents. For t-th pixel, the set of allocated topic labels is given by $z_t = \{z_{t*} \mid t^* \in \mathcal{T}\}$. The cluster label for t-th pixel might be defined as a function of the set $c_t = f(z_t)$, which is a major voting function in [3] or a spatial weighting function in [4, 5].

3. THE PROPOSED METHOD

Without loss of generality, a collection of documents $\underline{\mathcal{D}}$ could be generated for a given multi-spectral remote sensing image, if multiple reflectance values of a pixel could be regarded as a set of visual words for the pixel. If F spectral bands are used, there would be F collections of documents $\underline{\mathcal{D}} = \{\mathcal{D}_0, \mathcal{D}_1, \cdots, \mathcal{D}_{F\text{-}1}\}$.

3.1 Model

Given all of documents $\underline{\mathcal{D}} = \{\mathcal{D}_0, \mathcal{D}_1, \cdots, \mathcal{D}_{F\text{-}1}\}$, the uLDA assumes the following generative process as shown in Fig. 1 (b):

1) For each band, K topics are sampled from a Dirichlet prior $P(\Phi_k \mid \beta_k)$.
2) *Document Sampling*: For t-th pixel, a document index d_t is sampled from a prior distribution $P(d_t \mid \sigma, h)$.
3) *Topic Sampling*: Topic label for t-th pixel is sampled from a multinomial distribution *multinomial*(θ_i), where θ_i is mixture coefficient of the sampled d_t document.
4) *Word Sampling*: For each of F visual words, the attribute value of t-th pixel is sampled from a discrete distribution of topic z_t.

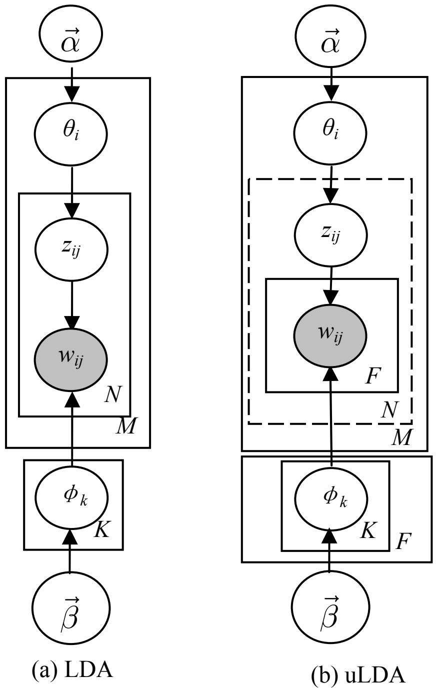


Fig. 1 Probabilistic graph model

3.2 Algorithm

The clustering algorithm using uLDA is given as follows:

Algorithm Gibbs_uLDA(**W**, α, β, K)
Input: word-count matrix **W**, parameters α, β, topic number K
Output: Topic matrix **Z**, multinomial parameters $\underline{\Phi}$ and Θ, estimated parameters α, β

// randomly initialization
Topic index matrix **Z**, document index matrix **D**
for all pixels $t \in \mathcal{T}$ **do**
Increment number of k-th topic in the document, which centers on t-th pixel: $n_t^{(k)} = n_t^{(k)} + 1$
End

// Gibbs sampling over burn-in period and sampling period
while not finished **do**
for all pixels $t \in \mathcal{T}$ **do**
// decrement number of w-th word associated with k-th topic on t-th pixel
for each band $f \in [1, F]$ **do** $n_{t,f}^{(w)} = n_{t,f}^{(w)}$ - 1 **end**

// decrement number of k-th topic in the document, which covers t-th pixel
for each document d_t **do** $n_{dt}^{(k)} = n_{dt}^{(k)}$ - 1 **end**

// multinomial sampling for document index

Sample document index $d_t \sim P(d_t \mid D_{t^-}, \mathbf{Z}, \mathbf{W})$

// Given document indexes, multinomial sampling
Sample topic index $z_t \sim P(z_t \mid \mathbf{Z}_{t^-}, D, \mathbf{W})$

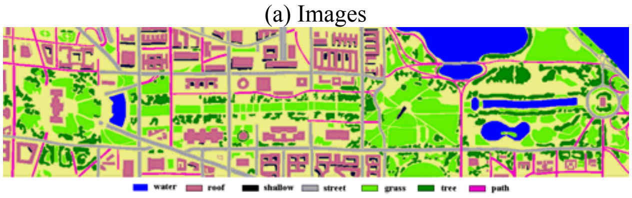
// for new assignment, increment counts and sums
for each band $f \in [1, F]$ **do** $n_{t,f}^{(w)} = n_{t,f}^{(w)} + 1$ **end**
for each document d_t **do** $n_{dt}^{(k)} = n_{dt}^{(k)} + 1$ **end**

// check converged and read out paramters
If converged and read out several sampls since last **then**
// the different parameters and read outs are averaged
read out parameter set $\underline{\Phi}$ and Θ
end
end

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental data

As shown in Fig. 2(a), a three-band false color image is used in our experiments, whose three channels are selected from a HYDICE image of Washington DC Mall with 191 bands. The number of selected three bands (i.e., red, green and blue) is 63, 52 and 36, respectively. There exist seven major geo-object classes, i.e., *water*, *roof*, *shallow*, *tree*, *grass*, *street* and *path*.


(a) Images


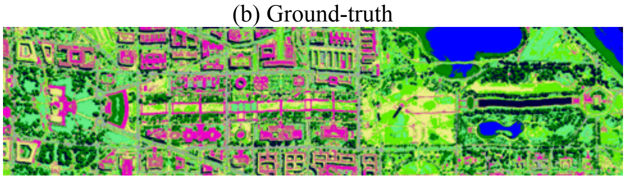
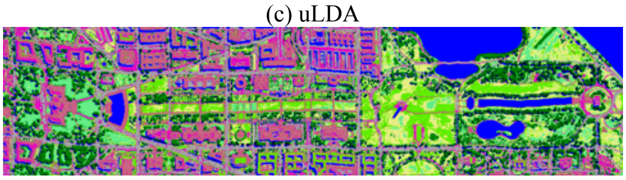
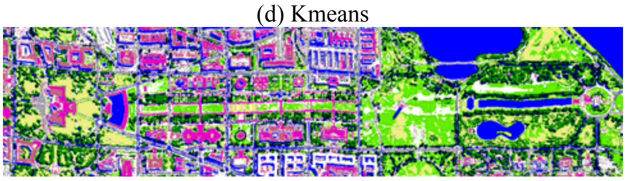
(b) Ground-truth

(c) uLDA

(d) Kmeans

(e) ISOdata

Fig. 2 Experimental data and results

It can be seen from Fig. 3 that (1) there exist seriously overlap between the histograms of the two geo-object classes *shallow* and *water* on each band; (2) the shape of the histograms is similar except the third band (i.e., blue band). This might be the reason why it is difficult to successfully separate the two geo-object classes during clustering.

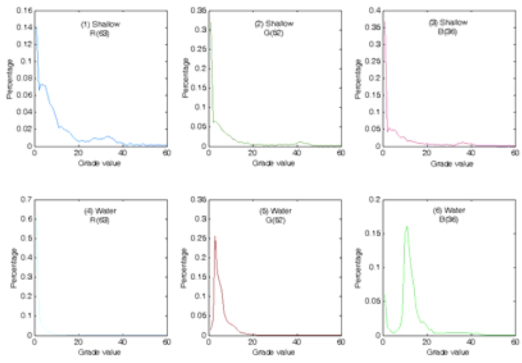


Fig. 3 Grade histogram of *shallow* and *water*

4.2 Experimental results

Three algorithms (i.e., uLDA, Kmeans and ISOdata) are used to clustering the false colour image shown in Fig. 2 (a). The clustering results are shown in Fig. 2 (c) ~ (e), respectively. It can be seen from Fig. 2(d) that nearly all of the *shallow* are incorrectly classified into the *water* by Kmeans. The potential reason might be both *water* and *shallow* are nearly be a same color. However, as shown in Fi.g.3 (c), *shallow* and *water* are well separated by the proposed method, although some *water* is also incorrectly classified as *shallow*.

4.3 Evaluation and Discussions

The generalized overall entropy defined (gOE) in [4] is used to quantitatively evaluate experimental results. Let h_{ck} be the number of pixels assigned to the cluster k within a ground-truth class c; $h_c = \sum_{k=1}^{K} h_{ck}$ be the total number of pixels within a ground-truth class c, and $h_k = \sum_{c=1}^{C} h_{ck}$ demote the number of pixels assigned to cluster k, where K is the number of clusters and C is the number of ground-truth class. On the one hand, the quality of a cluster is measured in terms of the homogeneity of the ground-truth classes within the cluster, i.e., the cluster entropy (CluE). For the k-th cluster, the cluster entropy E_k is given by

$$E_k = -\sum_{c=1}^{C} \frac{h_{ck}}{h_k} \log \frac{h_{ck}}{h_k} \tag{10}$$

On the other hand, given a ground-truth class c, the quality of a clustering result is measured in terms of the homogeneity of the cluster labels within the class, i.e., the class entropy (ClaE)

$$E_c = -\sum_{k=1}^{K} \frac{h_{ck}}{h_c} \log \frac{h_{ck}}{h_c} \tag{11}$$

The generalized overall entropy for a specific class c is defined as a linear combination of the class entropy E_c and the cluster entropy E_k

$$E_{generalized} = \beta E_c + (1-\beta) E_k \tag{12}$$

where $\beta \in [0,1]$ is a weight that balances the two measures [4] and β is set as 0.5 in our experiments; the cluster k covers the most pixels in the class c among all the clusters. Generally speaking, a smaller generalized overall entropy value indicates a higher homogeneity.

Table I. Quantitative evaluation of experimental results

Entropy / Class	ClaE		gOE	
	Kmeans	uLDA	Kmeans	uLDA
water	0.111561	0.552991	1.082	1.328
shallow	0.312971	0.508806		
roof	0.685607	0.795702		
street	0.530996	0.758737		
path	0.656281	0.803832		
tree	0.623385	0.735273		
grass	0.78519	0.832451		

Both the class entropy (ClaE) and the generalized overall entropy (gOE) are calculated and shown in Table I. Unfortunately, both the two evaluation indexes show that uLDA perform not as well as the Kmeans, although the *shallow* is well separated from the *water* in terms of qualitative visual inspection in Fig. 2 (c). The possible reason is that almost of *shallow* are not correctly allocated a same label as *water* by Kmeans. Meanwhile, a small fraction of *water* is labeled as *shallow* by the uLDA. Therefore, the cluster entropy of *shallow* in the Kmeans is relative lower than that in the uLDA.

5. CONCLUSION

In this paper, a novel framework is proposed to cluster multi-spectral remote sensing images. Although quantitative evaluation over the experimental results show that the proposed algorithm does not performs as well as traditional clustering algorithm, e.g., Kmeans, the proposed algorithm could well separate two geo-objects whose grade values are very similar to each other, e.g., the *shallow* and *water*. In the future, we are planning to enrich the proposed algorithm to achieve a better results.

ACKNOWLAGE

This work is supported by the National Basic Research Program of China (No. 2011CB707102), National Natural Science of China (No. 40901217; No. 41071259), and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] T. Hofmann, “Probabilistic latent semantic indexing,” in Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, 1999, pp. 50-57.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, Jan. 2003.

[3] M. Liénou, H. Maître, and M. Datcu, “Semantic annotation of satellite images using latent Dirichlet allocation,” *IEEE Geoscience and Remote Sensing Letter,* vol. 7, no. 1, pp. 28-32, Jan. 2010.

[4] W. Yi, H. Tang, and Y. Chen, “An object-oriented semantic clustering algorithm for high resolution remote sensing images using the aspect model,” *IEEE Geoscience and Remote Sensing Letter,* vol. 8, no. 3, pp. 522-526 May 2011.

[5] H. G. Akcay, and S. Aksoy, “Automatic detection of geospatial objects using multiple hierarchical segmentations ” *IEEE Transactions on Geoscience and Remote Sensing,* vol. 46, no. 7, pp. 2097-2111, July 2008.

[6] H. Tang, L. Shen, et al., “A MRF-based clustering algorithm for remote sensing images by using the latent Dirichlet allocation model”, Procedia Earth and Planetary Science 2 (2011) 358 – 363.

[7] X. Wang, and E. Grimson, “Spatial latent Dirichlet allocation,” in Proceedings of Neural Information Processing Systems Conference (NIPS) 2007, Vancouver, B.C., Canada, 2007, pp. 1577-1584.