A Semisupervised Latent Dirichlet Allocation Model for Object-Based Classification of VHR Panchromatic Satellite Images

Li Shen, Hong Tang, Yunhao Chen, Adu Gong, Jing Li, and Wenbin Yi

Abstract—Typically, object-based classification methods are learned using training samples with labels attached to image objects. In this letter, a semisupervised object-based method in the framework of topic modeling is proposed to classify very high resolution panchromatic satellite images using partially labeled pixels. In the stage of training, both topics and their co-occurred distributions are learned in an unsupervised manner from segmented satellite images. Meanwhile, unlabeled pixels are allocated user-provided geo-object class labels. In the stage of classification, each segment is classified as a user-provided geo-object class label with the maximum posterior probability. Experimental results show that the proposed method outperforms several SVM-based supervised classification methods in terms of both spatial consistency and semantic consistency.

Index Terms—Object-based image analysis, probabilistic topic models, semisupervised image classification.

I. INTRODUCTION

With the advancement of remote sensing technology, large collections of very high resolution (VHR) satellite images are becoming increasingly available to the public. An accompanying need is to effectively extract thematic information from VHR satellite images.

In order to achieve an effective classification of VHR satellite images, many classification approaches based on image segmentation have been proposed to take advantage of both spatial information and spectral information. The majority of these methods, e.g., object-based image analysis [1], typically regard segments as “image objects” and classify unlabeled objects using the classifier learned from a set of fully labeled objects. Although these methods can work well, there exist some problems in their practical applications: 1) training samples must be labeled according to the result of image segmentation, making it difficult to collect training samples of image objects in experimental fields before image analysis, and 2) individual ground truth points cannot be used directly to train object-based classifiers. Some methods, e.g., spectral–spatial methods [2], conduct pixel-based classification followed by spatial regularization using segmentation maps—where the segmentation is defined as an adaptive neighborhood for pixels and the informative statistical characteristics of image objects are not fully utilized. These observations motivate us to develop a new object-based semisupervised classifier, in which partially labeled pixels—instead of fully labeled image objects—are used to train an object-based classifier by using a probabilistic topic model. Benefiting from utilizing both labeled and unlabeled pixels in the stage of training [3], semisupervised learning has been proved to be effective in increasing classification accuracies in many remote sensing applications [4]–[7].

In recent years, probabilistic topic models, e.g., probabilistic latent semantic analysis [8] and latent Dirichlet allocation (LDA) [9]—which were originally proposed in text domain—have gained successful ground in natural image understanding [10]–[12], as well as satellite image analysis [13]–[17]. It is reported that the classification or clustering results obtained by topic models can rely more on semantic coherence. However, due to the unsupervised nature of basic topic models [18], the aforementioned methods based on topic modeling were previously developed for unsupervised image classification or clustering. Although much has been done to incorporate supervised information into topic models, most of the existing approaches are trained using labeled documents [19]–[23]. For this reason, they suffer from the same problem as the object-based methods mentioned previously—that is, their inability to utilize partially labeled pixels to train a classifier.

This letter presents a semisupervised object-based method to classify VHR panchromatic satellite images based on the LDA model [9], [24], which is referred to as the ssLDA. In the stage of training, both topics and their co-occurred distributions are learned in an unsupervised manner from segmented satellite images. Meanwhile, unlabeled pixels are allocated user-provided geo-object class labels. In the stage of classification, each segment is classified as a user-provided geo-object class label with the maximum posterior probability.

The remainder of this letter is organized as follows. In Section II, the proposed approach is presented in detail. Experimental results are given in Section III. Finally, the conclusion is drawn in Section IV.
TABLE I
QUANTITIES IN THE ssLDA MODEL

<table>
<thead>
<tr>
<th>QUANTITIES IN THE ssLDA MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
</tr>
<tr>
<td>$K$</td>
</tr>
<tr>
<td>$C$</td>
</tr>
<tr>
<td>$V$</td>
</tr>
<tr>
<td>$N_d$</td>
</tr>
<tr>
<td>$\tilde{\alpha}$</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td>$\eta$</td>
</tr>
<tr>
<td>$\phi_k$</td>
</tr>
<tr>
<td>$\phi_k$</td>
</tr>
<tr>
<td>$\pi_k$</td>
</tr>
<tr>
<td>$w$</td>
</tr>
<tr>
<td>$z$</td>
</tr>
<tr>
<td>$c$</td>
</tr>
<tr>
<td>$W$</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>$C$</td>
</tr>
</tbody>
</table>

II. METHODOLOGY

This section first defines correspondences of text-related terms in the image domain. Then, both the model and the algorithm are presented.

A. Mapping Image Into Text for the ssLDA Modeling

To introduce techniques used in the text domain to satellite images, it is necessary to build an analog of text-related terms—such as words, vocabulary, and documents—in image domain. For the proposed processing, the following analogs are defined.

1) Following the definition in [13] and [15], the grayscale value of a pixel is defined as a word.
2) The vocabulary is composed of the unique grayscale values of pixels.
3) Image segments, which are obtained by performing image segmentation on the original satellite image, are regarded as documents. Thus, a corpus of documents is equivalent to a satellite image. It should be noted that, to avoid destroying the boundaries and structures of geoclasses, the original satellite image is over-segmented [25].

To simplify the expression, a list of some related quantities in this letter is given in Table I.

B. Model

Compared with the graphical model of the LDA, the variable $c_{di}$ of geo-object class label for each word, as shown in Fig. 1(a), is added to model the one-to-one correspondence between pixels and geo-object class labels [9], [24]. The hashed fill pattern is used as “half-shaded” to denote that the class node is partially observed.

Given a set of segments originating from a panchromatic satellite image, the generative process of the ssLDA model is as follows.

1) For $k$th element of $K$ topics:
   a) Sample a topic specific grayscale distribution $\phi_k$ according to the Dirichlet distribution $[\phi_k \sim \text{Dirichlet}(\beta)]$.
   b) Sample a topic specific class distribution $\pi_k$ according to the Dirichlet distribution $[\pi_k \sim \text{Dirichlet}(\eta)]$.
2) For $d$th element of $D$ segments, sample a topic mixture proportion $\theta_d$ according to the Dirichlet distribution $[\theta_d \sim \text{Dirichlet}(\alpha)]$. Then, for the grayscale value $w_{di}$ and the class label $c_{di}$ of each pixel in $d$th segment, carry out the following sampling.
   a) Sample a topic $z_{di}$ according to the multinomial distribution over topics $[z_{di} \sim \text{Multinomial}(\theta_d)]$.
   b) Sample a grayscale value $w_{di}$ according to the multinomial distribution over grayscale values conditioned on topic $z_{di}$ $[w_{di} \sim \text{Multinomial}(\phi_{z_{di}})]$.
   c) If the class label $c_{di}$ is unobserved, it behaves according to the multinomial distribution over class labels conditioned on topic $z_{di}$ $[c_{di} \sim \text{Multinomial}(\pi_{z_{di}})]$. If not, the sampling is ignored, and $c_{di}$ takes the original value of the given training pixel.

C. Algorithm

As shown in Fig. 1(b), Gibbs sampling is used to conduct model inference. For ease of exposition, assume that, if $i$th pixel in $d$th segment is indexed by $s = (d, i)$, then the expressions of $w_{s}$, $z_{s}$, and $c_{s}$ are equivalent to that of $w_{di}$, $z_{di}$, and $c_{di}$, respectively. In addition, the subscript $-s$ denotes a quantity that excludes data on the site $s$. The algorithm of the proposed model can be summarized in the following three steps.

1) Step 1: Preprocessing. In this step, the parameters in the model are initialized, and the satellite image is over-segmented. For the proposed model, symmetric priors are used for all Dirichlet hyperparameters. Thus, as shown in Fig. 1(a), there are six scalars that need to be set. These include the number of segments $D$, the number of topics $K$, the number of classes $C$, and the Dirichlet hyperparameters $\alpha$, $\beta$, and $\eta$. In addition, two matrices with the same size as the original panchromatic satellite image, i.e., matrices of topics $Z$ and class labels $C$, are initialized by assigning random topics for all pixels and random class labels to the unlabeled pixels. Furthermore, a segmentation map is created by over-segmenting the original satellite image to constitute a corpus of documents.

2) Step 2: Training. Topics are learned in an unsupervised manner from the segmented satellite images. As shown in Fig. 1(a), it can be found that the generative process
Fig. 1. Graphic model and algorithm of the ssLDA. The shaded nodes represent observed random variables, and the hashed fill node denotes partially observed random variable class. The intuitional explanations of the model are shown at the end of the red and dashed arrows. (a) Model. (b) Algorithm.

of a class label is almost the same as that of a word (i.e., grayscale value of the pixel). Analogous to the deduction in [26], for each site \( s \), the conditional posterior of the topic \( z_s \) can be given by

\[
P(z_s = k \mid W, Z_{-s}, C) \propto \frac{n_{w_s k, -s} + \beta}{\sum_{v' = 1}^{V} n_{w_s v', -s} + V \beta} \cdot \frac{n_{c_s k, -s} + \eta}{\sum_{c' = 1}^{C} n_{c_s c', -s} + C \eta} \cdot \frac{n_{d_s k, -s} + \alpha}{\sum_{k' = 1}^{K} n_{d_s k', -s} + K \alpha}
\]

where \( n_{w_s k, -s} \) and \( n_{c_s k, -s} \) are the numbers of times the grayscale value \( w_{di} \) and the class label \( c_{di} \) occur with \( k \)th topic with the exception of the value on site \( s \), respectively. Likewise, \( n_{d_s k, -s} \) refers to the number of times \( k \)th topic occurs in \( d \)th segment with the exception of the value on site \( s \).

Meanwhile, unlabeled pixels are allocated user-provided class labels. Specifically, for each site \( s \), the conditional posterior of the class \( c_s \) can be given by

\[
P(c_s = c \mid W, Z_{-s}) \propto \frac{n_{c k, -s} + \eta}{\sum_{c' = 1}^{C} n_{c c', -s} + C \eta}
\]

It is worth noting here that the class label of a pixel should be kept untouched if the pixel corresponds to the labeled training pixel. Otherwise, the pixel will be allocated a new class label by sampling using the aforementioned equation.

The sampling process of the topic \( z_s \) and class \( c_s \) is repeated until convergence or until the maximum number of iterations has been reached. The final sampling results are used to approximate the parameters of document topic mixing proportion and per-topic class distribution

\[
\theta_{d,k} = \frac{n_{k d} + \alpha}{\sum_{k' = 1}^{K} n_{k' d} + K \alpha}
\]

\[
\pi_{k,c} = \frac{n_{c k} + \eta}{\sum_{c' = 1}^{C} n_{c' k} + C \eta}
\]

The topic mixing proportion reflects co-occurrence relationships of topics, and the topic specific class distribution reflects dependence between latent topics and user-provided geo-object class labels.

3) Step 3: Classification. Each segment is classified as a user-provided geo-object class label with the maximum posterior probability

\[
c^* = \arg \max_{1 \leq c \leq C} P(c \mid d) = \arg \max_{1 \leq c \leq C} \left( \sum_{k = 1}^{K} \pi_{k,c} \cdot \theta_{d,k} \right)
\]

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Data

As shown in Fig. 2(a), a panchromatic QuickBird image of a suburban area with a size of 900 × 900 pixels and 0.6-m spatial resolution was used in the experiments. Fig. 2(b) shows the ground truth map, which includes six geo-object classes of interest, i.e., building, road, shadow, water, tree, and field. The segmentation map was obtained to constitute a corpus of documents using the entropy rate superpixel segmentation algorithm, which has been proven to be both effective and efficient in [25]. The optimal number of segments (i.e., documents) \( D = 2000 \) was experimentally derived.
B. Comparison With Existing Methods

To evaluate the effectiveness of the proposed method, the performance of the ssLDA has been compared with that of three state-of-the-art classification methods: 1) the pixel-based SVM; 2) the spectral–spatial SVM—where pixel-based SVM classification was followed by a majority voting within the adaptive neighborhoods defined by image segmentation (termed as SVM+MV) [2]; and 3) the segment-based SVM, where each region or segment was treated as a basic processing unit and the attribute of a segment was represented by the average grayscale value of all pixels within the segment (termed as Seg-SVM). The latter two methods are based on image segmentation.

The ssLDA initialized the Dirichlet priors as symmetric priors, i.e., $\alpha = 50/K$, $\eta = 50/C$, and $\beta = 100$. The number of classes was set to six according to the distribution of geo-object classes, and the number of topics was set to ten by fivefold cross validation. For the three SVM-based methods, a multiclass one-versus-one SVM classification with RBF kernel by means of LIBSVM was used [27]. The optimal parameters $C'$ and $\gamma$ were determined by fivefold cross validation. For all methods based on image segmentation, the same segmentation map was adopted.

Table II gives the global (overall accuracies as well as kappa coefficients) and class-specific (producer accuracies) classification accuracies for different methods. The corresponding classification maps are shown in Fig. 2(c)–(f), respectively, where each geoclass is represented by a color.

From visual inspection, all methods based on image segmentation achieve a more compact and smoother classification result when compared to the pixel-based SVM. Therefore, the advantages of performing image segmentation by enforcing spatial consistency over the classification are confirmed. Furthermore, due to a certain degree of spectral overlap between road and building, they are seriously misclassified with each other in the classification map of the Seg-SVM method. Similar misclassification can also be observed for tree and field in the classification map of the SVM+MV method. The phenomena might be explained by the fact that, although both the Seg-SVM and the SVM+MV embed the spatial context information from the segments, they can only ensure spatial consistency and lack a mechanism to distinguish between different geo-objects with similar spectral characteristics.

However, the proposed ssLDA benefits from the topic modeling and works differently. The co-occurrence relationship of different geo-objects can be modeled by the mixture of topics. Therefore, when these geo-objects are observed frequently in a different scene, the mixture proportion features can be utilized to differentiate them. Table II shows that the ssLDA method yields the best global as well as a majority of the best class-specific classification accuracies. In particular, the producer accuracies for building, field, road, and tree are significantly improved. In general, the proposed ssLDA method can obtain a better classification result in terms of both spatial consistency and semantic consistency.

C. Influence of Different Sizes of Labeled Pixels

As the number of labeled training pixels may affect the classification results, it is therefore worth exploring how the performance of the proposed method behaves with different settings on the size of labeled pixels. A set of experiments with
different proportions of the ground truth pixels for training, i.e., 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, and 40%, is conducted. Labeled training pixels for each class were acquired at random from the ground truth, and the remaining labeled pixels of ground truth were used for testing. The same labeled pixels were used to fit the SVM+MV, which was used as a baseline method. It should be noted that, in the ssLDA, all of the unlabeled pixels are also used for training.

Fig. 3 shows the overall accuracies of both the ssLDA and the SVM+MV against the proportions of labeled training pixels. As can be seen, there are two obvious observations: 1) across all different sizes of labeled pixels, the ssLDA method achieves a better performance than the SVM+MV, and 2) the performance difference between the ssLDA and the SVM+MV has correlation with the size condition of labeled pixels, i.e., when the size of labeled pixels is relatively small, these two methods perform comparably. In contrast, when the size of labeled pixels is large, the ssLDA significantly outperforms the SVM+MV. This is due to the reason that SVM-based methods are not relatively sensitive to the size of labeled training pixels [28]. By comparison, both unlabeled and labeled pixels are utilized in the mechanism of the proposed ssLDA, and with the increase in size of labeled pixels, the model enforces stronger constraints that align learned topics with user-provided geo-object classes, i.e., class labels of labeled training pixels. Thus, in the proposed method, the classification accuracy can be improved considerably when a large number of labeled training pixels are available. It should be noted that, although the requirement of large sizes of labeled training pixels may sometimes be difficult to achieve in practical applications, the proposed ssLDA method provides a promising solution that fully utilizes the labeled training pixels to improve classification accuracy.

IV. CONCLUSION

In this letter, a semisupervised object-based method has been proposed to address the problem of VHR panchromatic satellite image classification. The method extends the LDA model by incorporating the information of partially labeled pixels for training. Segments obtained by image segmentation are used as documents to enforce spatial regularization over the classification results. Experimental results show that the proposed approach is superior to three aforementioned SVM-based methods. Furthermore, this approach is not tied to a specific segmentation algorithm. Any method that can produce a reasonable oversegmentation of satellite images may meet the requirement of the proposed method.

REFERENCES