SEMANTIC ANALYSIS OF SATELLITE IMAGE TIME SERIES

Teodor Costachioiu[1], Rodica Constantinescu[1], Bashar AlZenk[1] and Mihai Datcu[2]

1) Politehnica University of Bucharest
2) DLR - German Aerospace Center, Oberpfaffenhoffen, Gemany

ABSTRACT

Large archives of satellite images have been created over time. The existence of these archives enables us to extract evolutions of the same of same geographic area over time, creating satellite image time series (SITS). As SITS represent an amount of information far greater than individual images, their analysis is complex and difficult. In this paper we propose a new unsupervised SITS analysis method based on the latent Dirichlet allocation (LDA) model, a hierarchical model originally developed for text analysis. In this model documents are represented as random mixture of latent topics, each topic being characterized by a distribution over words. This paper extends the use of LDA model for satellite image time series analysis by proposing a description language for SITS modeling according to the LDA model, and is applied on a SITS of 11 Landsat TM scenes acquired in 2007.

Index Terms— SITS, satellite image time series, Latent Dirichlet Allocation, unsupervised classification

1. INTRODUCTION

The development of modern imaging sensor technologies and the ongoing of the current remote sensing mission over the course of time have led to creation of large satellite image archives. This archives enables us to extract observations of the same geographical area at different moments in time, allowing the creation of Satellite Image Time Series (SITS). SITS have a far greater complexity than single satellite images, embedding spatial and temporal information. While supervised analysis method [1] can provide good results, with high user-semantic meaning, they depend on the size and quality of the training set, which is difficult and time-consuming to create. Often, in case of historical imagery, it is not possible to obtain the ground-truth information required for the creation on test and training data. In this situation we can rely on unsupervised methods [2],[3],[4]. Unsupervised SITS analysis methods can be applied in situations where no ground-truth in available, but are less efficient than supervised methods. The main problem in unsupervised SITS analysis in the "semantic gap", or is the divergence between the low-level features that can be extracted the data and the "knowledge" specific for each user and application.

In this paper we propose a new approach to SITS analysis that aims to extract dynamic evolution classes with a high user-semantic meaning. The method proposes a new SITS encoding method according to the "bag-of-words" model used in text domain, followed by a semantic analysis performed with Latent Dirichlet Allocation algorithm [5]. The paper is organized as follows. In the next section, we will briefly describe the LDA model for text analysis. Then, Section 3 details the way we exploit the LDA model for the semantic annotation of satellite image time series, by defining a new method of encoding the time series according to the "bag-of-words" model, required by LDA. Experiments and results are presented in Section 5, followed by a conclusions section.

2. LATENT DIRICHLET ALLOCATION

In the effort to bridge the semantic gap, some algorithms originally developed for text analysis have been used for model-based analysis of remotely sensed data [6],[7]. Most of the text analysis methods are based on the "bag-of-words" text model, in which only the word count is considered, the order of the words in a document being ignored. Starting from this assumption, documents are modeled as mixtures of (hidden) topics. Among these algorithms we can mention statistical text modeling tools such as pLSA [8] and generative probabilistic models such as Latent Dirichlet Allocation (LDA) [9]. The pLSA model brings a probabilistic approach, representing topics as probability distributions over the vocabulary space, documents being characterized by probability distribution over the topic space.

LDA model can be considered as an extension of pLSA, treating the topic mixtures parameters as variables drawn from a Dirichlet distribution, each topic being characterized by a multinomial distribution over the word space. As such, LDA is a complete generative model and, unlike pLSA, the LDA model has the ability to generalize over new documents that are not present in the training set.

In text domain we can define the *words* as basic units of discrete data. All words are extracted from a vocabulary indexed by $\{1 \dots V\}$. Words can be represented

as unit-basis vectors that have a single component equal to one and all other components equal to zero. Considering a vocabulary of size V the vth word in the vocabulary is represented by a V-vector w such that $w^v = 1$ and $w^u = 0$ for $u \neq v$.

A *document* is a sequence of N words denoted by $w = (w_1, w_2, ..., w_N)$ where w_n is the nth word in the sequence.

A *corpus* is a collection of M documents denoted by $D = \{w_1, w_2, ..., w_M\}$.

Considering the above definitions the LDA generative process can be described as follows:

1. Choose $N \sim Poisson(\xi)$
2. Choose $\theta \sim Dir(\alpha)$
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim Multinomial(\theta)$
   (b) Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

The Poisson distribution used to model document length is not critical, document length distributions can be used as needed.

The dimensionality k of the Dirichlet distribution corresponds to the total number of topics, and is presumed known. The mixture of topics is given by:

$$\text{Dir}(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} ... \theta_k^{\alpha_k - 1} \quad (1)$$

where $\alpha = \{\alpha_1 ... \alpha_k\}$ are the parameters of the Dirichlet distribution and Γ is the gamma function.

The word probabilities are parameterized by the $k \times V$ matrix β, with $\beta_{ij} = p(w^j = 1)$.

Considering LDA model parameters α and β, the the marginal distribution of a document is given by:

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^{N} p(w_n|\theta, \beta)\right) d\theta \quad (2)$$

Starting from a collections of documents, in the LDA training phase we estimate the model parameters α and β such that the log-likelihood of the entire collection is maximized. However, the relation (2) is not tractable for exact inference: the solution is to use approximate inference algorithms such as Laplace approximation, Gibbs sampling or variational expectation maximisation. In this work we use the original method proposed by Blei *et.al*, which employs variational inference.

Regardless of the approximation method chosen, the number of topics k is presumed known and is the only user-given parameter of the LDA model.

Often there are situations when the number of topics is unknown. In such cases we can estimate the number of topics by computing the *perplexity* of a set of test documents. In text modeling perplexity is used to measure of the ability of a model to generalize to unseen data. Considering a set of M documents, $D_{test} = \{w_1, w_2 ... w_M\}$,

each document containing a number of words N_d, perplexity can be defined as:

$$perplexity(D_{test}) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right) \quad (3)$$

A lower value of the perplexity indicates a better LDA model. When the number of topics is unknown several LDA models can be trained for different number of topics, the model characterized by the lowest value of perplexity induicating the optimal topic number.

3. SEMANTIC ANALYSYIS OF SATELLITE IMAGE TIME SERIES

To extent the LDA towards SITS analysis the first step is to define a description language for satellite image time series. The main purpose of this language is to convert the satellite image time series - which are ordered sequences of events - into a "bag-of-words" representation. On the other hand, in the original LDA model topics are represented as multinomial distributions over the vocabulary space, with a number of degrees of freedom equal with the number of words. To apply the LDA model in SITS analysis, the "words" of this description language must be positive, integer numbers.

The proposed method follows the framework presented in fig. 1. An unsupervised clustering performed by k-means is applied on each scene, extracting a fixed number of classes. Then, to the classes obtained by clustering the order of each scene is attached and as a result we obtain the words such as $\langle SITS_{order}; cluster\_class \rangle$. This way, words contain both temporal and spectral information, and the order of the words can be ignored.
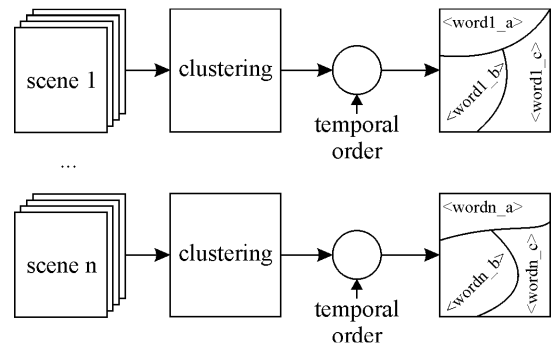

Fig. 1 The proposed word generation method for SITS semantic analysis

By the way words are created, each word becomes an unique feature to the image that generates it. This has several advantages: it can be applied directly on the data in the archives, eliminates the need for radiometric corrections; the actual class number assigned by k-means has no importance – it is the LDA algorithm that finds semantic meaning for the above defined words. The proposed method has only a few parameters: the number of the k-means classes and the number of LDA topics.

4. EXPERIMENTS AND RESULTS

Experiments were performed on a SITS containing 11 Landsat TM scenes covering the area of Bucharest and its surroundings, acquired in 2007, with the acquisition dates given in table 1. The size of the studied area 1100 x 1500px, respectively 33 x 45 km. The area contains several large land-cover and land-use classes, including urban areas, forests, agricultural land and water bodies.

TABLE 1
ACQUISITION DATES FOR BUCHAREST SITS

16/05/2007	26/06/2007	28/07/2007	29/08/2007
25/05/2007	03/07/2007	04/08/2007	14/09/2007
10/06/2007	19/07/2007	20/08/2007	

Following the proposed method, in the experimental work we sought to evaluate the effect of the number of k-means classes and the optimum number of topics. We performed clustering in the spectral domain of each scene for a number of $c = 5, 10, 20$ and 50 k-means classes. As the studied image time series contains 11 scenes, the total number of words is 55, 110, 220 and 550 words.

After generating these words we have created documents at pixel level, a document containing all words that describe the spectro-temporal evolutions of the pixel.

As LDA is a generative model, we exploit its ability to generalize over unseen data and we use only 10% of the original data set for training. To find the optimal number of topics we have trained LDA models for $3 \div 70$ topics and we have plotted the perplexity evolutions, according to fig. 2 and fig. 3.
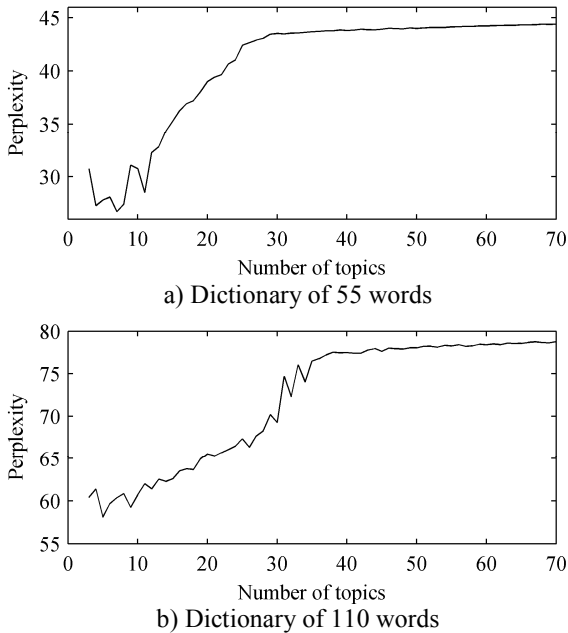
a) Dictionary of 55 words

b) Dictionary of 110 words

Fig. 2 Perplexity evolutions as a function of topic numbers, evaluated for different dictionary sizes.

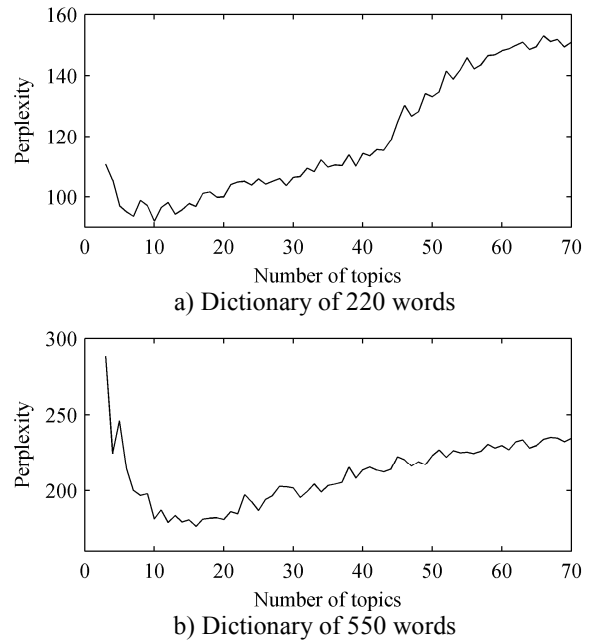a) Dictionary of 220 words

b) Dictionary of 550 words

Fig. 3 Perplexity evolutions as a function of topic numbers, evaluated for different dictionary sizes. The lowest value of perplexity indicates the optimum number of topics

By examining the perplexity evolutions we notice that increasing the number of word increases the overall perplexity of the LDA models, which is to be expected as a higher number of words means a higher complexity of the LDA model. By comparing the perplexity evolutions we find that as the number of words is increased a higher number of dynamic evolutions in indicated by the perplexity. In our case this is obtained for a vocabulary of 550 words, in which according to the perplexity graph in fig. 3 b) we can extract a number of 16 LDA topics, as presented in fig. 4.
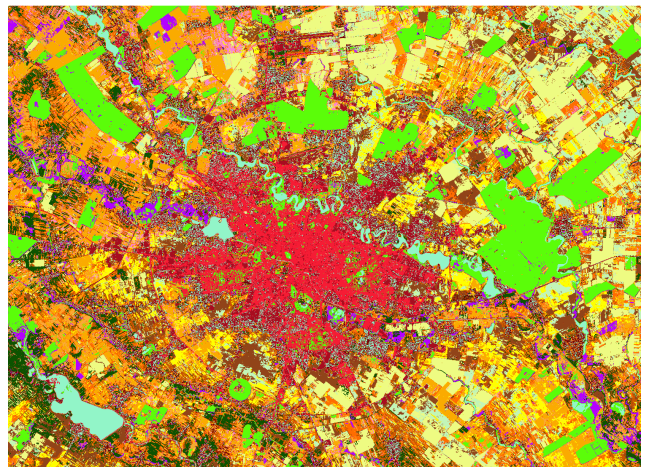
Fig. 4 Dynamic evolution classes extracted by the proposed method for a 550 words dictionary

Due to inexistence of appropriate ground truth information an exact evaluation of the proposed method is not possible. However, we can compare the results of the classification in fig. 4 against land use and land cover classes extracted from Urban Morphological Zones and CORINE Land Cover datasets from 2006 [10], with the results being displayed in table II.

TABLE II

COMPARISSON BETWEEN THE CLASSES EXTRACTED BY THE PROPOSED METHOD AND KNOWN LAND USE AND LAND COVER CLASSES

	Urban area	Forests	Water bodies	Agriculture
LDA C1	74.11	0.44	0.17	25.27
LDA C2	14.80	10.24	1.75	73.21
LDA C3	4.16	0.25	0.08	95.51
LDA C4	6.13	0.21	0.10	93.56
LDA C5	67.61	0.61	0.47	31.31
LDA C6	13.04	0.88	0.16	85.92
LDA C7	5.54	1.09	0.27	93.10
LDA C8	2.48	0.35	0.07	97.09
LDA C9	51.57	0.42	28.84	19.17
LDA C10	0.85	0.07	0.05	99.02
LDA C11	90.09	0.37	1.22	8.33
LDA C12	1.83	0.12	0.04	98.01
LDA C13	26.00	3.25	0.39	70.36
LDA C14	13.40	0.65	0.32	85.63
LDA C15	12.64	70.11	2.14	15.11
LDA C16	10.29	0.39	0.15	89.18

By comparing the results in table II we notice the extraction of classes that show a strong membership to one of the known land use / land cover classes. Such is the case of agriculture (classes LDA C10, LDA C12, LDA C8, LDA C3 and LDA C4) and urban areas such as class LDA C11. Forests are also identified by class LDA C15. By running at pixel level the proposed method has also identified as forests some patches of vegetation in the urban area (parks, etc.) The lack of detailed ground truth information does not allow a precise evaluation of this class.

There are also some temporal dynamic classes that span over one or more land use types. This is the case of LDA C9 class, which covers both urban and water bodies, the LDA C1, LDA C2, LDA C5 and LDA C13 classes which describe temporal evolutions from both urban and agricultural areas.

5. CONCLUSION

In this paper we have presented a new method for unsupervised analysis of satellite image time series which, with a proper SITS-to-words encoding, allows the use of semantic analysis tools from text domain. The method can be applied at pixel level, preserving the spatial resolution of the original time series. By comparing the experimental results with known land use and land cover classes we have demonstrated that the proposed method extracts classes of temporal evolutions with a high relevance for the user.

As the experiments have shown that the number of extracted classes is dependent on the quantization level of each scene, future work will include testing of other quantization methods that are able to extract the optimal number of clusters from each scene.

6. REFERENCES

[1] Costăchioiu, T.; Lazărescu, V.; Datcu, M., "Classification of scene evolution patterns from Satellite Image Time Series based on spectro-temporal signatures," Signals, Circuits and Systems (ISSCS), 2011 10th International Symposium on, vol., no., pp.1-4

[2] T. Häme, I. Heiler, J.S. Miguel-Ayanz, "An unsupervised change detection and recognition system for forestry," International Journal of Remote Sensing, Vol. 19, 6, p. 1079-1099, 1998

[3] Heas, P.; Datcu, M.; , "Modeling trajectory of dynamic clusters in image time-series for spatio-temporal reasoning," Geoscience and Remote Sensing, IEEE Transactions on , vol.43, no.7, pp. 1635- 1647, July 2005

[4] Julea, A.; Méger, N.; Bolon, P.; Rigotti, C.; Doin, M.-P.; Lasserre, C.; Trouvé, E.; Lăzărescu, V.N.; , "Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns," Geoscience and Remote Sensing, IEEE Transactions on , vol.49, no.4, pp.1417-1430, April 2011

[5] D. Blei and M. Jordan. Modeling annotated data. Technical Report UCB//CSD-02-1202, U.C. Berkeley Computer Science Division, 2002.

[6] Lienou, M.; Maitre, H.; Datcu, M.; , "Semantic Annotation of Satellite Images Using Latent Dirichlet Allocation," Geoscience and Remote Sensing Letters, IEEE , vol.7, no.1, pp.28-32, Jan. 2010

[7] Bratasanu, D.; Nedelcu, I.; Datcu, M.; , "Bridging the Semantic Gap for Satellite Image Annotation and Automatic Mapping Applications," Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol.4, no.1, pp.193-204, March 2011

[8] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," J. Mach. Learn. Res., vol. 42, no. 1/2, pp. 177–196, Jan. 2001.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, no. 5, pp. 993–1022, Mar. 2003.

[10] European Environment Agency, http://www.eea.europa.eu