A UNIFIED UNSUPERVISED CLUSTERING ALGORITHM AND ITS FIRST APPLICATION TO LANDCOVER CLASSIFICATION

Yong Yu, Isabelle Bloch *

Département TSI, CNRS URA 820
Ecole Nationale Supérieure des Télécommunications
46 rue Barrault
75013 Paris - FRANCE
{yuyong, isabelle.bloch}@tsi.enst.fr

Alain Trouvé

LAGA, CNRS UMR 7539
Institut Galilee
Université Paris 13
Av, Jean-Baptiste Clement
93439 Villetaneuse
trouve@math.univ-paris13.fr

ABSTRACT

The problem of classification is so fundamental that it has been intensively investigated by many researchers from different domains. In this paper, we present a novel unsupervised clustering algorithm derived from the techniques of probabilistic modeling which is implemented by a stochastic gradient algorithm. Then its application to challenging landcover classification based on Daedalus data of the SMART project is explored by combining both spectral feature and spatial contextual information. Our first experiments show its potential usefulness in remote sensing.

1. INTRODUCTION

The problem of classification is so fundamental in pattern recognition that it has been intensively investigated by many scientists and engineers from different domains such as computer science, mathematical statistics, psychology, etc. As a result, many classification algorithms have been developed from the different points of view of applications and theories (see [1] and references therein). A lot of artificial vision tasks of different levels such as image segmentation, image grouping and content based image retrieval etc. can be directly or convertibly expressed as the problems of unsupervised and supervised methods of data classification. Since unsupervised clustering methods, unlike their supervised counterparts, do not need an off-line labeling procedure for the selected training set which is either time-consuming to implement or unstable to guarantee the accuracy of labeling results, such methods have been preferably explored in our case of landcover classification. Recently, the so-called pairwise grouping techniques have been largely investigated for the task of data classification [2, 3, 4] which can be described as: firstly, define the (dis)similarity measure between data pairs which are the proper extracted feature representation according to the application. Thus a graph structure is obtained from an affinity matrix defined on this measure and data pairs; then apply some grouping criterion and optimization method among all the possible clustering of all the elements which usually fall into linear discrimination analysis. The global structures can be detected by the dynamical behavior of the local pairwise interaction.

In this paper, we will focus on the clustering method in the same spirit as the previous mentioned approaches and based on a novel optimization criterion from the techniques of probabilistic modeling. Then we have applied this unsupervised clustering algorithm to landcover classification problem on Daedalus data. The optimization criterion and corresponding clustering algorithm are given in section 2. One of our contribution is that our method unifies both spectral feature and spatial contextual information in the same framework: Some implementation detail is given in section 3 for spectral feature and in section 4 for combining spatial contextual information. Finally some preliminary landcover classification results are shown in section 5 before the short summary.

2. PROBABILISTIC MODEL AND CLUSTERING ALGORITHM

For all the stochastic approaches (discrete models), we should give some detail about probabilistic model which is applied to the problems (landcover classification tasks in our case). An image I is considered as the realization of an independent sampling process from a given probability measure μ on an abstract space $\mathcal{X}$, in other word, the image I is defined a family of iid r.v. $I = (I_i)_{1 \leq i \leq |I|}$.

Now, any partition $C = (C_1, \cdots, C_K)$ of I with cluster

*Prepared through collaborative participation in the Information Society Technologies - Humanitarian Demining Programme (SMART IST-200-25044) sponsored by European Committee.

number K known a priori, is explained as the implementation of some image grouping process by a similarity criterion associated to $\mathcal{X}$. More precisely, one class C_u in any partition C can be defined a sequence of Bernoulli variables in $\mathcal{X}$.

$$X \doteq (X_x)_{x\in\mathcal{X}} = \begin{cases} 1 & \text{if } x \in C_u \\ -1 & \text{else} \end{cases}$$

For any subset A of I, we can defined

$$X_A = \inf_{x\in A} X_x,$$

so that X_A should be interpreted as a sign function taking the value 1 when A is included in C_u and -1 otherwise.

Assume now we would like to label I into K disjoint sets, or clusters, $C = (C_1, \cdots, C_K)$ so that, in each cluster, two elements of I are more likely to belong simultaneously to the same cluster. In order to select a meaningful criterion, we consider the following game: "Consider a given clustering C, choose uniformly in I a data indexed by i, then select uniformly in the same cluster other than I_i a new data indexed by j. What is the probability that both data I_i and I_j belong simultaneously to the same cluster or to its complement ?". From the previous game, we can define, for each clustering C of I, a score $S_N(C)$, which is the computed probability. This number is a performance index for a clustering in grouping data according to a given probability measure defined on Bernoulli variables $(X_x)_{x\in\mathcal{X}}$.

By direct computation, we get $S_N(C)$ as following:

$$S_N(C) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{K}\left(\frac{1}{|C_k|}\sum_{i,j\in C_k} P(X_{I_i} = X_{I_j} \mid I)\right) \quad (1)$$

More properties about $S_N(C)$ can be found in [5] which explores mainly its connection with the common K-means algorithm. In our implementation of the unsupervised clustering, we have adopted a hierarchical structure of binary partition tree: recursively partition each node from root node (which represents the whole data set) into two sub-clusters. To make the algorithm more clear, we give some outline of the computation procedure in each node of the current layer in the tree:

1. Initialize the binary partition by random selection and compute its initial S_N.

2. repartition the data set by flipping the cluster label of a random singleton k. If

$$\delta W \doteq \frac{W_0'}{|C_0|+1} + \frac{W_1'}{|C_1|-1} - \frac{W_0}{|C_0|} - \frac{W_1}{|C_1|}$$

where $W_l = \sum_{i,j\in C_l} P_{ij}$, $W_l' = W_l + sgn(k \in C_l)\sum_j P_{kj}$, $l \in \{0, 1\}$, is negative, then restore the k's labeling.

3. iterate the previous step until S_N reaches maximization.

The unsupervised clustering algorithm we have proposed depends on: the joint probability distribution $P(i, j)$ and the layer number of the partition tree. The former can be from the pairwise distances defined on I , which is feature oriented; the latter decides the number of the final clustering: in our experiment, we have always chosen $K = 2^p, p \in N$. If in some case, K is chosen arbitrarily, then the binary tree is firstly built with a number of leaves larger than but closer to it, then the merging procedure is applied to the smallest clusters. The total computation complexity is $o(|I|^2)$. Therefore a compromise should be made between speed and accuracy of the algorithm to deal with large dimension of remote sensing images which is explained in the following sections.

3. SPECTRAL FEATURE BASED CLASSIFICATION

We have applied the clustering algorithm directly to the spectral signature based classification on the multi-spectral Daedalus data. That is, the partition has been done in the spectral signature space where each pixel I_i is represented by a vector $(I_{i1}, \cdots, I_{iL})^T$ (L is the channel number). To calculate the joint probability distribution, for the moment, we assume that the different channels of data give independent response. i.e.: $P(i, j) = \prod_{l=1}^{L} P(I_{il} = I_{jl})$. Furthermore, we assume that for each channel of data, the joint probability is:

$$P(I_{il} = I_{jl}) = \exp(-\frac{|I_{il} - I_{jl}|^2}{\sigma_l^2}) \quad (2)$$

where σ_l, the standard deviation is estimated from the available data set.

In order to overcome the huge computation on Daedalus data of large size, the partition on the spectral signature space is approximated by a small set of pixels (64*64) sampling uniformly from I. Since the empirical distribution based on the pixel set with such a size is well approximated to the ideal one, we still maintain the validation of the partition while reducing the dimension of computation.

Then for the most data not appearing in the sampling set, they are labeled by a simple KNN decision rule, i.e., the label of each unknown pixel is appointed by the major voting on the labels of its k (k=20 in our case) nearest neighbors.

4. SPATIAL FEATURE COMBINATION

In general, introducing the spatial-contextual information is helpful to improve the accuracy of the classification[6, 7], e.g., to reduce the noise interference. We have explored it again in our unsupervised clustering framework.

The first trial is to consider the joint probability $P(i,j)$ is the simple composition of two kinds of independent sources: one still from the spectral signature denoted by $P_{ss}(i,j)$ while the other comes from their relative spatial position denoted by $P_{sp}(i,j)$. For $P_{sp}(i,j)$ we have used the same probabilistic model as $P_{ss}(i,j)$. This kind of idea appears in many energy based image segmentation techniques. However, within our clustering framework, the observed results show that no matter how the parameter is tuned in $P_{sp}(i,j)$, the classification result is always in the plate like form and not satisfying. One explanation is that we have emphasized the spatial factor too much.

The idea here is to give more factor for $P_{ss}(i,j)$ than $P_{sp}(i,j)$. We have adopted a non-linear term as:

$$W(i,j) = P_{ss}(i,j)(1 + \lambda 1_{d_{ss}(i,j) \leq th^*} P_{sp}(i,j)) \qquad (3)$$

where λ is some tuning parameter (λ =1 in our case), $d_{ss}(i,j)$ is a distance metric defined on the spectral signature space and th^* is a threshold option (the estimated standard deviation is adopted in our case).

Now $W(i,j)$ cannot be explained as some probability distribution term and it seems that our general clustering cannot be applied directly. However, the meaning of $W(i,j)$ is clear: the more reward is obtained from the spatial information besides from the spectral signature only if data pair is not too dissimilar. In short, $W(i,j)$ is explained as a reward term combining both the spatial contextual information and spectral signature quantity. If $W(i,j)$ replaces $P(i,j)$ in Equ. 1, then $S_N(C)$ is still a similarity criterion based on $W(i,j)$. There is no additional changement for our clustering algorithm.

For the same reason to reduce the computation complexity, we have firstly divided the whole image into small regular windows (32*32) before applying the clustering process. As a result, a post-processing step is needed to merge the clustering results into several large segmentation regions depending on the statistics for the clusters. For the moment, we have simply computed the empirical mean and standard deviation for each clusters then applied the Lloyd's version of K-means algorithm [8] to agglomerate the clusters to the given number of classes.

5. EXPERIMENT ON LANDCOVER CLASSIFICATION OF DAEDALUS DATA

We have done some experiments of landcover classification on the Daedalus data (12 channels from visible blue to thermal infrared, its image of 4th channel is shown in Fig. 2(a)) at the test site Glinska Poljana, Croatia, which is crucial to our project mission of minefield level 1 survey. Our data collected from sensors cannot offer the information such as the mine positions themselves as that in [9] which can be used to generate directly the "danger" map. Therefore, in our case, such a map will be produced mainly by change detection results from the data of different time and the knowledge of potential danger degrees for different ground truth: i.e., the different status of the vegetation area (abandoned or used), the boundary of the forest region, etc. Therefore, it is interesting to know whether our clustering algorithm can describe accurately our target classes, e.g. the forest region.

Here only the results of combining the spatial contextual information are presented for the sake of space limit since we have observed that they are always better than those using only spectral features. Fig.1 is the small size (512*512) result of test site Glinska Poljana with class number $K = 4$. We also display the corresponding ground truth map given by Mrs. R. Pernar of CROMAC in Fig. 1(c) to validate the output of our algorithm in Fig. 1(b). Although the ground truth map gives more detail description (it has 19 classes), the detection result by clustering algorithm gives more accurate boundaries for our classes of interest, such as bushes and vegetation regions.The classification result of the whole test site with $K = 8$ is displayed in Fig.2(b). It should be remarked that forest region has been enhanced largely due to combining spatial contextual information, which is more interesting to our target mission.

6. SUMMARY

We have introduced a new clustering algorithm derived from the techniques of probabilistic modeling giving the intuitive partition meaning. Furthermore, it can be flexibly extended to a common framework embeding the spatial information. The preliminary application to the landcover classification shows its usefulness in the remote sensing domain. Our current efforts are focused on both developing the methods to extract the linear structures (road, river, railway, etc) in our Daedalus data under the same probabilistic model and comparing our result with those of other algorithms, e.g., MultiSpec©.

7. ACKNOWLEDGMENTS

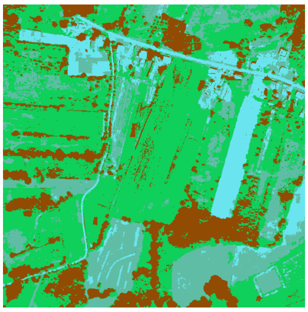
The authors from ENST thank the support from EC supported SMART project and also all the project partners for their kindly help and discussion, especially DLR for providing all the test images.

8. REFERENCES

[1] A. K. Jain, P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

(a)

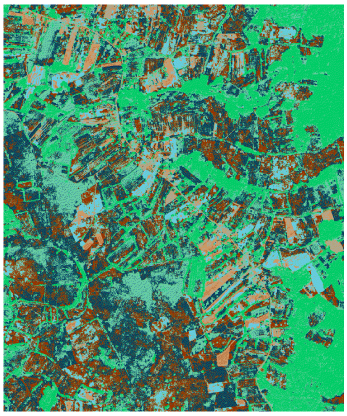
(b)



(c)

Fig. 1. Unsupervised clustering results on daedalus data of the selected smaller region : (a) original image (4th Channel with size 512*512); (b) ground truth map drawn by Mrs. R. Pernar (This map is a courtesy of CROMAC and Mrs. R. Pernar) ; (c) output of the algorithm (4 classes) .

(a)



(b)

Fig. 2. (a) Channel 4 daedalus data on the test region with size 2448*2048 ; (b) Unsupervised clustering results on the selected test combining spatial contextual information with 8 classes output.

[2] Thomas Hofmann and Joachim M. Buhmann, "Pairwise data clustering by deterministic annealing," *IEEE Trans. PAMI*, vol. 19, no. 1, pp. 1–14, January 1997.

[3] Shi J. and Malik J., "Normalized cuts and image segmentation," in *Proceeding of IEEE Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 731–737.

[4] Y. Gdalyahu, D. Weinshall, and M. Werman, "Self organization in vision: stochastic clustering for image segmentation, perceptual grouping, and image database organization," *IEEE Trans. on PAMI*, vol. 23, no. 10, pp. 1053–1074, October 2001.

[5] Yu Yong and Trouvé Alain, "A non-linear k-means algorithm and its application to unsupervised clustering," preprint, 2002.

[6] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[7] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, 1985.

[8] S. P. Lloyd, "Least squares quantization in pcm," *IEEE transactions on Information Theory*, vol. 28, pp. 129–137, 1982.

[9] Irene YU-Hua Gu and Tardi Tjahjadi, "Detecting and locating landmine fields from vehicle- and air-borne measured ir images," *Pattern Recognition*, vol. 35, pp. 3001–3014, 2002.