CONSENSUAL CLUSTERING FOR LAND COVER MAPPING

Marine Campedel, Ivan Kyrgyzov

TELECOM ParisTech, Institut MINES-TELECOM, CNRS LTCI
46 rue Barrault, 75013 Paris, France

ABSTRACT

In this article we propose to illustrate the ability of consensual clustering to provide mining tools in the context of land cover unsupervised classification. The proposed algorithm is based on individual co-association matrices related to several input clusterings that are combined using a Mean Shift optimization procedure. This provides valuable clusters in terms of interpretation and also information about the data to be clustered, which could be useful to discriminate between easily classified pixels and the other ones, requiring human expertise. The interest of our approach is demonstrated using the Boumerdes dataset provided by SERTIT and CNES, in the context of the 2003 earthquake.

***Index Terms*—** land cover classification, consensual clustering, mining tool

1. INTRODUCTION

Our first motivation is to provide interpreters with mature and scientifically evaluated tools. This evaluation takes place in realistic applicative context like rapid mapping and needs rich and interpreted datasets. Thanks to collaboration with SERTIT and CNES, valuable datasets can be used to demonstrate the ability of tools in case of major disasters like earthquake, flooding or fires.

In this article we propose to evaluate the consensual clustering proposed by Kyrgyzov [1] in 2008 (but never published) for land cover classification, using Boumerdes (Algeria) dataset [2]. Consensual clustering is an intuitive and unsupervised way to combine different clusterings on the considered data: this can be useful when nothing is known about the best classifiers to be used, the parameters to be tuned or even the features to be considered. In other words it is a way to mimic the human behavior when discovering a new scene. Consensual clustering is, at first, a mining tool helping interpreters to tune their automatic algorithms. The produced consensual classification is not only an extra data clustering, it is also a way to compare all input clusterings and to provide information about the considered data in terms of ability to be classified. The idea is to discriminate between data that can automatically be classified and data that need more attention from the interpreter. Hence consensual clustering can also be used as an evaluation tool.

In this article we propose three main ideas. In the following section we first theoretically describe the Mean Shift-based consensual clustering, and propose three stability measures at cluster and pixels levels. We then define an experimental protocol involving a new dataset provided by CNES and SERTIT. We finally discuss the mining and evaluation abilities of our algorithm in the context of Boumerdes earthquake land cover mapping.

Thanks to CNES and SERTIT for their useful datasets.

2. THEORETICAL BASIS

2.1. State of the art

Classification combination algorithms are not new in the literature: Condorcet popularized voting strategies in the 18th century. But it can be noticed that if the literature is rich when dealing with known target classes [3], it regained interest in the mid 2000 concerning clusterings.

The idea of clustering is to group data according to a given similarity measure for further data exploration or retrieval. In our case, data are geographical locations or pixels, often characterized by spectral or/and textural features in images. There are many ways to extract these features, to compare them and then to cluster them. Different clusterings provide different information that can be redundant or complementary and the combination of all these results can produce a valuable clustering as well as interesting clues about the induced choices. The problem is then to define a valuable combination criterion while avoiding a too high complexity.

Lots of studies have been proposed in the literature. Our approach is directly related to the definition of co-association matrices and to their combination, as proposed in [4, 5]. This matrix, denoted A^p can be defined for any clustering p and the general term for two clustered data u and v is:

$$A^p_{uv} = \begin{cases} 1, & \text{if } u \text{ and } v \text{ are in the same cluster,} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For P considered clusterings, the average co-association matrix A is computed as:

$$A = \frac{1}{P}\sum_{p=1}^{P} A^p \quad (2)$$

For large P, A_{uv} estimates the probability for two data u and v to belong to the same cluster, according to the input clusterings. The classical question is now to find the consensual (binary) co-association matrix D that will best approximate the average co-association matrix A.

2.2. Proposed algorithm

The criterion proposed in [6] consists in minimizing the square error between D and A, with N the number of clustered data:

$$E(D) = \|D - A\|_F^2 = \sum_{u=1}^{N}\sum_{v=1}^{N}(D_{uv} - A_{uv})^2 \quad (3)$$

Equation (3) can be rewritten as:

$$E(D) = \sum_{u=1}^{N}\sum_{v=1}^{N} D_{uv}(1 - 2A_{uv}) + \sum_{u=1}^{N}\sum_{v=1}^{N} A_{uv}^2 \quad (4)$$

The first term in Eq.(4) has no square degree because D is a binary matrix. In addition, the last term in Eq.(4) is a constant and does not influence the minimization of error E. The quadratic objective function Eq.(4) may be solved exactly for small data sets using efficient methods, in contrast to the optimization of *NMI* criterion in [4]. We propose an original way to perform this optimization, based on Mean Shift estimation.

The P input clusterings with J^p clusters are now considered as P binary allocation matrices B^p Eq.(5), where $p \in 1, ..., P$.

$$B_{uj}^p = \begin{cases} 1, & \text{if sample } u \in \text{cluster } j, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The matrices are concatenated into a single matrix **B** and the new feature space is $\{0,1\}^d$, where $d = \sum_{p=1}^{P} J^p$. Each line b_u of this concatenated matrix describes the allocation of each data in all input clusters (presented in columns).

The samples $\{b_u\}$ are located on a hyper circle, since they simultaneously satisfy a hyper plane equation $\sum_{j=1}^{d} b_{uj} = P$ and a hyper sphere equation $\sum_{j=1}^{d} b_{uj}^2 = P$. Therefore vectors $\{b_u\}$ may be normalized by a constant $\sqrt{P}$ such that their square norm is 1. Hence we get the relationship between the average co-association A and b_u:

$$A_{uv} = b_u b_v' \quad (6)$$

Moreover let μ_j be the centroid of the consensual cluster C_j, $\mu_j = \sum_{v \in C_j} b_v / n_j$, $v \in C_j$. The square norm of μ_j is:

$$\|\mu_j\|^2 = \frac{1}{n_j^2}\left\|\sum_{v \in C_j} b_v\right\|^2 = \sum_{u \in C_j}\sum_{v \in C_j} A_{uv}/n_j^2 \quad (7)$$

	Initialize $j = 1, c_i = 0$ for $i = 1, ..., N$
Step1	Choose u unlabeled, i.e. $c_u = 0$, Else stop and return c (consensual labels)
Step2	Initialize the local mean estimation $k = 1$, $\mu_k = b_u$, $r_k = 1$, with n_k the number of data inside the window $W(\mu_k, r_k)$ centered on μ_k with radius r_k,
Step3	Shift the local mean: $\mu_{k+1} = \frac{1}{n_k}\sum_{b_i \in W(\mu_k, r_k)} b_i$, $r_{k+1} = \|\mu_{k+1}\|$ Iterate Step2 and 3 until convergence (denoted as $_{conv}$): $\mu_j = \mu_{conv}$, $r_j = r_{conv}$
Step4	Assign u to the estimated mean μ_j Assign $c_i = j, \forall i : \|b_i - \mu_j\|^2 < r_j^2$, $j = j+1$, Goto Step1.

Table 1. Proposed algorithm, called MSCC (Mean Shift Consensual Clustering)

The square error E proposed in Eq. (4) is then reduced to its left term:

$$\begin{aligned} E^2(J, \{C_j\}_{j=1}^J) &= \sum_{j=1}^{J}\sum_{u \in C_j}\sum_{v \in C_j}(1 - 2A_{uv}) \\ &= \textstyle\sum_{j=1}^{J} n_j^2 (1 - \frac{2}{n_j^2}\sum_{u \in C_j}\sum_{v \in C_j} A_{uv}) \\ &= \textstyle\sum_{j=1}^{J} n_j^2 (1 - 2\|\mu_j\|^2) \end{aligned} \quad (8)$$

with $\{C_j\}_{j=1}^J$ the consensual clusters and n_j the number of elements in C_j. A global minimum of the error E^2 in Eq. (8) is reached when maximizing the norms of local mean vectors $\|\mu_j\|^2 > 0.5$ while jointly maximizing the number of samples n_j in consensual clusters. This problem can be seen as a well known non parametric density estimation by mean shift vectors [7], whose modes are the consensual clusters means μ_j. We propose in [1] to use the multivariate Epanechnikov kernel [8], well adapted to our binary vectors b_u since global convergence is proven is this case [9].

The final algorithm is proposed in Table 1. The process begins with each data b_u being a cluster and the idea is to agglomerate data in order to minimize the proposed energy. Lets consider one state of our algorithm: J clusters are already formed and N_J data are still considered as unitary clusters (corresponding to an energy -1). To merge one of these data b_u into an existing cluster μ_j, this data has to be in the neighborhood of μ_j and the energy must decrease i.e. $E_{after}^2 < E_{before}^2$, with $E_{before}^2 = E_0^2 + E_{C_j}^2 + E_{b_u}^2$, $E_{after}^2 = E_0^2 + E_{C_j U b_u}^2$ and E_0^2 the common energy corresponding to the data untouched at the current step. When developing these energies, it comes

$$\mu_j b_u' > 0.5 <=> \|b_u - \mu_j\|^2 \leq \|\mu_j\|^2 = r_j^2 \quad (9)$$

2.3. Related stability scores

The presented algorithm, called MSCC (Mean Shift Consensual Clustering) provides a new clustering of the data, based on the non-parametric estimation of all local means μ_j. We propose to derive 3 measures to help the data mining task, based on Eq. (7). The data-based stability measure is:

$$S_u^k = \frac{1}{n_k} \sum_{i \in C_k} A_{ui}, \text{with } u \in C_k \tag{10}$$

When taking the mean over a cluster:

$$S^k = \frac{1}{n_k} \sum_{u \in C_k} S_u^k = \|\mu_k\|^2 \tag{11}$$

We also derive a criterion between two clusters:

$$S^{kl} = \frac{1}{n_k} \frac{1}{n_l} \sum_{u \in C_k} \sum_{v \in C_l} A_{uv} \tag{12}$$

These criteria are used to exploit both the consensual clustering and the input clusterings in order to derive information on the produced clusters (how much consensual they are), their relationships to other clusters as well as clues about the easiness to classify specific data. This is illustrated in the following experiments.

2.4. Complexity

The complexity of the presented algorithm is $O(N^2)$ and cannot directly be applied to a whole scene. However efficient implementations can be obtained at scene level: we do not detail this point in this paper and present only local result exactly corresponding to the presented algorithm.

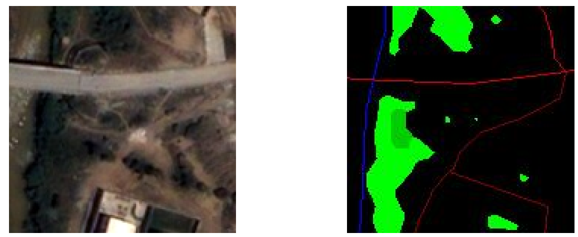
3. EXPERIMENTS

3.1. Boumerdes Dataset

As already mentioned, we benefit from a very interesting dataset including QuickBird images (60cm, multispectral) before and after the earthquake at Boumerdes, as well as several vector data describing the scene. We are interested in land cover classes like water areas (*sea, lake, river*), *vegetation* (high and low), *beach*, as well as structural classes like *railway* and different types of *roads*. Information about buildings and refugees camps are also provided but we do not use it here.

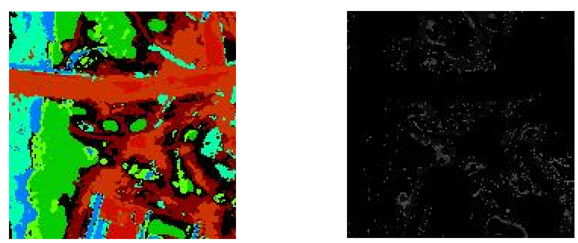
3.2. Experimental conditions

In the illustrative example we use one sample (cf Fig. 1(a)) of the image immediately after the earthquake, with size 150×150, resulting in $N = 22500$ pixels. We performed 3 K-Means clustering of the whole image (2301×3334 pixels) using Monteverdi tool [1] directly on the pixel values in the 4 spectral bands, with $K = 10, 15, 20$. We focus our example on one small portion of the scene, illustrated on Fig. 1(b), containing 5 different classes of interest: 2 *vegetation* classes (high and low) appearing in green, *river* in blue, 2 *road* classes in red. Note that the two *road* classes and the *river* are represented as linear objects, since we do not have thickness values. Black pixels are classified as *bare soil* in the dataset but this is not clearly true since they also correspond to parts of roads, river as well as buildings, we call them *unclassified* in the following.



(a) (b)

Fig. 1. (a) Studied sample (2003-05-23, 3 colors composition) and (b) Projected ground truth: the linear objects do not completely reflect reality, since we do not have thickness values.



(a) Classified image (b) Pixel instability ($1 - S_u^k$)

Fig. 2. Results provided by the consensual clustering.

3.3. Clusters validity

In this example, 12 clusters have been identified by the consensual clustering. The stability of each cluster is very high (over 0.9 and several times equal to 1) except for one cluster (cluster 8). This means that the 3 input K-Means group data similarly in this sample, except for data in cluster 8. We provide the consensually clustered image (cf Fig. 2(a)): the colors associated to the 12 clusters are derived from the ground truth colors, using the saturation and value of the mostly represented class in the cluster and the weighted hue according to class effectives. We roughly observe that we can detect the 3 main classes *river*, *vegetation* and *road* but the discrimination

[1] Orfeo Toolbox - Monteverdi - http://orfeo-toolbox.org

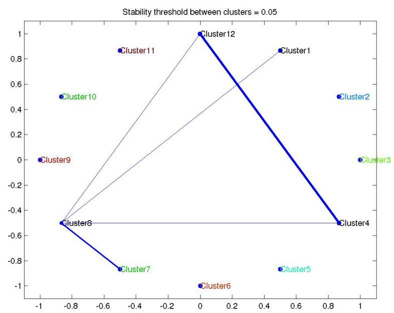


Fig. 3. Cluster relationships (S^{kl}).

between the two types of *vegetation* or the two types of *roads* cannot be performed. We also observe shadows pixels in the cluster containing mainly *river* pixels. However, on the previously unclassified pixels we can get relevant extension of the vegetation areas as well as roads that were not indicated in the ground truth. Note that the high stability scores indicate that the input clustering were behaving the same. We can conclude that these input clusterings are relevant to make the main classes emerge from the scene but not sufficient to discriminate between more specific classes. More work should be done to say if it is because of the K values or because of the input features.

It is interesting to note that 4 of the 12 consensual clusters are purely black, containing only unclassified pixels, with high stability scores. Using our inter-cluster stability scores S^{kl} (cf Eq. (12)) we can produce a connected graph, whose connection thickness is proportional to the obtained scores. We only show connections higher than 0.05 (this means that at least 5% of pixels are shared by the related clusters, according to the input clusterings), in Fig. 3. It appears that these black clusters (1, 4, 8 and 12) have connections. The less stable cluster is the one logically having the higher number of connections (cluster 8). In the spatial domain, this corresponds to neighbouring pixels between different objects.

3.4. Data stability

In Figure 2(b), we propose to show the estimated instability at each pixel ($1 - S_u^k$). This can be considered as a way to estimate the easiness of each pixel to be classified by different algorithms. If the score is close to 0 (appearing in black), this means that the consensus between input clusterings is very high; it appears in white for unstable data, i.e. data that are classified with different neighbors according to the input clusterings. We observe, consistently with the preceding observations on clusters, that most pixels are very stable. The unstable ones seem to be located at the frontier of objects, in the dark areas of the consensually classified image. Again, it would indicate to the interpreters that these pixels (corresponding to the previously mentioned *black* clusters) deserve specific attention since they reflect the differences between the input clusterings in delimitating objects.

4. CONCLUSION

In this article we proposed a new algorithm to compute consensual clustering, based on Mean Shift density estimation and co-association matrices. We also derived 3 stability measures to help the mining and interpretation process. We finally proposed an illustrative example on the Boumerdes dataset to explain the interest of our proposal. It demonstrated the ability of the consensual clustering to identify classes of interest and also to focus attention on less stable data, that are not classified the same way by the input clusterings. This methodology can also be applied to the whole scene, using the three images in the dataset; the resulting clusters are then different, putting in evidence areas evolving in time.

5. REFERENCES

[1] I. Kyrgyzov, *Recherche dans les bases satellitaires des paysages et application au milieu rubain : clustering, consensus et catgorisation*, Ph.D. thesis, Telecom ParisTech, 2008.

[2] ©copyright SERTIT distribution CNES, "Dataset boumerdes," 2009.

[3] L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley-Interscience, 2004.

[4] A.L.N. Fred and A.K. Jain, "Combining multiple clusterings using evidence accumulation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 835–850, 2005.

[5] A.K. Jain A. Topchy and W. Punch, "A mixture model for clustering ensembles," in *in Proc. SIAM Conf. on Data Mining*, 2004, pp. 379–390.

[6] P. Michaud and F. Marcotorchino, "Modèles d'optimisation en analyse des donnés relationnelles," *Mathématiques et Sciences Humaines*, vol. 67, pp. 7–38, 1979.

[7] K. Fukunaga and L.D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IT*, vol. 21, no. 1, pp. 32–40, 1975.

[8] V. A. Epanechnikov, "Non parametric estimation of a multivariate probability density," *Theory Prob. Appl. (USSR)*, vol. 14, pp. 153–158, 1969.

[9] V. Ramesh D. Comaniciu, "Real-time tracking of nonrigid objects using mean shift," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 142–149, 2003.