International Journal of Soft Computing 1(2) : 137-142, 2006

New Fuzzy Clustering Algorithm Applied to RMN Image Segmentation

Nabila Ferahta, Abdelouahab Moussaoui, Khier Benmahammed and Victor Chen
Department of Technology Science, University Ferhat-Abbas of Setif-Algeria

Abstract: An entirely automatic procedure for the classification of cerebral tissues from Magnetic Resonance Nuclear imaging (MRN) 3D of the head are described in this study. This procedure doesn't make any assumption nor on the number of classes nor on the shape of the density. Indeed, this last is estimated by a non parametric method, it is about the method of the Parzen's Kernel. A new objective function is proposed to improve the FCM algorithm by the addition of one term of entropy aiming to maximize the number of good ordering. A supplementary correction is operated by a probabilistic procedure said of fuzzy relaxation including the probabilities of the neighboring points. The validation of the algorithm is made on simulated data and on real cerebral imaging RMN.

Key words: Clustering, automatic classification, SKIZ, markov fields, image segmentation, Maximum Posterior Marginal (MPM)

INTRODUCTION

The segmentation can be described as the definition of groupings, in the parametric space, where the points are associated to the different sets of values of similar intensities in the different images. As consequence, in this approach, the process of grouping is the main step in the procedure of segmentation[1,2,3] and the techniques based on the automatic grouping are reputed to be more robust for the separation of the different tissues in presence of noise and imprecise data, in relation to the techniques of contours detection[4].

Besides, the uncertainty is largely present in the medical images because of the noise (during the acquirement) and of the effects of partial volumes. It means that the values of the voxels, especially to the borders between volumes of interest, correspond to the miscellanies of different anatomical tissues, because of the low resolution of the sensors. As consequence, the borders between tissues are not defined correctly and the adherences in the limits of the regions are intrinsically fuzzy. To shortcoming all these considerations, our choice for the segmentation of the anatomical tissues carried itself on the methods of analysis of data whose principle is founded mainly on the fuzzy automatic grouping.

The first studied fuzzy method concerns the algorithm of the Fuzzy C-Means (FCM) by J. Bezdek[5,6]. The FCM algorithm requires a priori definition of the number of classes and its results depend of this number. The application of this algorithm to the segmentation of medical images is described in[7,8]. The second method exposed in this article is an improvement of the first aiming to optimize the number of classes and to maximize the number of good orderings and it by the addition of one regulating term server to minimize the entropy of the histogram of the classified image. Another refinement is operated by an algorithm that permits to take in consideration the spatial relation between the different voxels, it is about the fuzzy relaxation that is an extension of the probabilistic relaxation[9].

ESTIMATION OF PROBABILITY DENSITY FUNCTION (PDF)

The data to classify are in a $\mathbb{R}^n$ space of dimension. Let's consider a sample of N data $(X_i)_{1\le i\le N}$ represented by N points with: $(X_i = \chi_{i,1}\ \chi_{i,2},\ \ldots,\ \chi_{i,n})$. The probability density function is defined on $\mathbb{R}^n$ by a function f satisfying at least to the following conditions:

$$\forall x \in \mathbb{R}^n, f(x) \ge 0 \ \ et \ \int_{\mathbb{R}^n} f(x)\ dx = 1 \qquad (1)$$

In the practice, this function f is unknown; several methods exist to estimate it. The most widespread method consists in approaching the function f by a histogram, but the gotten evaluation doesn't present the properties of continuity required by most applications. Besides, the widths of the classes (or elements) of the histogram are delicate to determine.

Among the methods of evaluation of the density by a continuous function, one of the most known is the method of the k nearest neighbors. The estimation of the probability density function is defined by:

Corresponding Author: Nabila Ferahta, Depattment of Technology Science, University Ferhat-Abbas of Setif-Algeria

$$\hat{f}(x) = \frac{k-1}{2Nd_k(x)} \quad (2)$$

where k is the fixed number of the nearest neighbors (typically $k = \sqrt{N}$ [10]) and d_k (x) is the distance of x to its k^{th} nearest neighbor (while classifying by order growing the distances of x to each of the N given data, d_k (x) is the K^{th} distances).

The estimation by the nearest neighbor's method doesn't permit to get a derivable function and besides it will give the peaks corresponding to data situated in the densest regions. In the literature, one considers that these continuous estimations but non derivable of the density are not smoothed enough and remain approximate enough[10,11].

The based methods solely on the maximum likelihood permit to derive the estimated function, but the interpolation enters the data is merely heuristic. The verisimilitude of the estimated density function $\hat{f}$ for the sample $(X_i)_{1\le i\le N}$ is defined by:

$$L\left(\hat{f}/(X_i)_{1\le i\le N}\right) = \prod_{i=1}^{N} \hat{f}(x_i) \quad (3)$$

Maximizing the verisimilitude doesn't permit to estimate the density outside of the sample $(X_i)_{1\le i\le N}$ $(X_i)_{1\le i\le N}$ (i.e. $\hat{f}$ (x) for x ≠ X_i). It is necessary to impose some restrictions to $\hat{f}$ permitting an acceptable interpolation of $\hat{f}$ enters the data. These restrictions or penalties have for goal to smooth the estimated function. In dimension 1, one uses the most often a criterion of the following type:

$$\ell_\alpha(f) = \sum_{i=1}^{N} \log f(x_i) - \alpha \int_{-\infty}^{\infty} (f'')^2 \quad (4)$$

The first term

$$\sum_{i=1}^{N} \log f(x_i)$$

is the logarithm of the verisimilitude that it is necessary to maximize, the second term

$$\alpha \int_{-\infty}^{\infty} (f'')^2$$

is a penalty that it is necessary to minimize and that is controlled by the parameter α. This parameter is adjusted to smooth more or less the estimation of the density[10,11]. In dimension 2 or more, this method of the penalties is more delicate to put in work, it is necessary to find the good compromised between the parameter of smoothing and the verisimilitude. In this type of approach, the part of heuristic is important. Because of this required lack, it seemed preferable to us the classical kernel's method to estimate the density (even named Parzen-Rosenblatt's[1] method) described below.

A. parzen-rosenblatt's method: In Parzen-Rosenblatt's method, we estimate the density of probability while using a convolution kernel. The kernel is a function k that is generally itself a function of probability density. In this description, we take like kernel the multi-normal function (centered and reduced) definite by[11,12]:

$$\forall x \in \mathbb{R}^n, \quad \hat{f}(x) = \frac{1}{Nh^n} \sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \quad (5)$$

h being the parameter of smoothing of the estimation.

In this density approach, each data X_i contribute in the same way to the calculation of $\hat{f}$ and this contribution depends on h. The kernel k being a unimodal and positive function, the contribution of every data to $\hat{f}$ be added and is worth at more

$$\frac{1}{Nh^n}$$

(the maximal contribution of a data X_i to $\hat{f}$ (x) is gotten when x = X_i). This estimation of the probability density corresponds to a convolution of the function k with the function definite by:

$$\Delta(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_i(x)$$
$$\text{with } \delta_i(x) = \begin{cases} 0 & \text{if } x = X_i \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

The parameter h corresponds to the square root of variance of the kernel k. More h, will be small, more the kernel will be narrow and $\hat{f}$ will present some peaks of probabilities to the points X_i. In this setting, h be called the window of the estimation or window of smoothing. This type of estimation of the probability density function depends on the choice of the smoothing window[12].

FCM CLASSIFICATION

A. definition of the classical criteria: The stage of classification consists in minimizing the criteria defined by the Eq. 7 that is not anything else that a generalization of the classic criteria of the k-middle (sum of the intra-class distances):

$$J = \sum_{i=1}^{C} \sum_{J=1}^{N} \left(u_{ij}\right)^m d^2\left(x_j, c_i\right) \qquad (7)$$

under the following constraint:

$$\sum_{i=1}^{C} u_{ij} = 1 \quad \forall j$$

where m characterizes the degree of fuzziness (when m grows we introduce more fuzzy), C represents the number of classes, N numbers of its pixels, c_i is here the characteristic vector of the class i and d^2 (x_j, c_i) is the Euclidian distance (it is possible to use another distance[6] between the point j and the prototype of the class).

The minimization of J takes place in two stages: in the first stage we minimize the functional in relation to the; c_i then the c_i being fixed, we minimize the functional in relation to the u_{ij}[5,13].

The algorithm is represented by (one supposes that C and m are well known):

$$\begin{cases} u_{ij} = \dfrac{1}{\displaystyle\sum_{k=1}^{C} \left(\dfrac{d^2\left(x_j, c_i\right)}{d^2\left(x_j, c_k\right)} \right)^{\frac{1}{m-1}}} & si \quad I_j = \phi \\ \begin{cases} u_{ij} = 0 & i \notin I_j \\ \sum_{i \in I_j} u_{ij} = 1 & i \in I_j \end{cases} & si \quad I_j \neq \phi \end{cases} \qquad (8)$$

$$with \;\; I_j = \left\{ i / 1 \le i \le C, d^2\left(x_j, c_i\right) = 0 \right\}$$

B.-definition of a new criteria: It appeared interesting to us a given algorithm in which we would not have to fix the number of classes but that would determine an optimal number of classes automatically. In this optics, we were inspired by works achieved by two teams. The first have been done by H. Frigui and R. Krishnapuram[14]. The characteristic function to minimize includes two terms. The first is a classic term characterizing the intra-class distance (generalized to the fuzzy case) and corresponds to the Eq. 7. The second term aims to maximize the number of good points in every class. It is about a pondered sum of adherence degrees.

The other team[15,16] worked also on the topic but remains in the non fuzzy case. The goal was to minimize the entropy of the histogram of the classified image. The second term of the criteria is therefore the next one:

$$J_2 = -\alpha \sum_{i=1}^{C} p_i \log\left(p_i\right) \qquad (9)$$

where p_i is a prior probability of the class i.

In our case we spread the previous works to the fuzzy case. The criterion that we minimize is the next one:

$$J = \sum_{i=1}^{C} \sum_{J=1}^{N} \left(u_{ij}\right)^m d^2\left(x_j, c_i\right) - \alpha \sum_{i=1}^{C} p_i \log\left(p_i\right) \qquad (10)$$

under the constraint:

$$\sum_{i=1}^{C} u_{ij} = 1 \quad \forall j \in \left[0, N\right]$$

In the whole continuation we take m = 2. While considering that u_{ij} represents the probability for the i pixel to belong to the class j, we define the probability of the class i as being:

$$p_i = \frac{1}{N} \sum_{j=1}^{N} u_{ij} \qquad (11)$$

The update of the parameters is gotten while minimizing the criteria alternately then in relation to the prototype c_i in relation to the u_{ij}. At every iteration, we keep the classes whose probability is superior to a certain doorstep only. In the fuzzy case[16,17], only the classes of non hopeless probability are kept.

$$c_i = \frac{\sum_{j=1}^{N} u_{ij}^2 x_j}{\sum_{j=1}^{N} u_{ij}^2} \qquad (12)$$

$$u_{ij} = A + B \quad where$$

$$A = \frac{\dfrac{1}{d^2\left(x_j, c_i\right)}}{\displaystyle\sum_{k=1}^{C} \frac{1}{d^2\left(x_j, c_k\right)}}$$

$$B = \frac{\alpha}{2Nd^2\left(x_j, c_i\right)} \left(1 + \log\left(p_i\right) - \frac{\displaystyle\sum_{k=1}^{C} \frac{1 + \log\left(p_k\right)}{d^2\left(x_j, c_k\right)}}{\displaystyle\sum_{k=1}^{C} \frac{1}{d^2\left(x_j, c_k\right)}} \right) \qquad (13)$$

In[14], H. Frigui and R. Krishnapurams have decreased the parameter α. In this manner, their algorithm finds a partition in a number of classes that is close to the optimum number of classes since the first iterations. We choose the initial value of α in such a way that the two terms of the criteria J (Eq. 10) are in the same way then we decrease the functional like the following manner:

$$\alpha(k) = \alpha(0)\exp\left(-\frac{k}{\tau}\right)\frac{\sum_{i=1}^{C}\sum_{J=1}^{N} u_{ij}^2 d^2(x_j, c_i)}{\sum_{i=1}^{C} p_i \log(p_i)} \qquad (14)$$

In this manner, once the number of optimum classes reaches, the classification is not biased by the term of entropy. In the followed tests, that parameter τ is fixed to 30 and α (0)to 2. K is the number of iterations.

FUZZY RELAXATION

A major inconvenience of the methods based on the automatic grouping is that they completely disregard information contained implicitly in the spatial coordinates of the pixels: the probability that the neighboring pixels belong to the same class is bigger than when the pixels are distant. For it, we considered coordinated them spatial of the pixels like a supplementary characteristic in the vector of attributes.

We consider here the probabilistic relaxation procedure, an extension of the labeling process by relaxation[9]. The starting point of the procedure is the situation where, after a certain numbers of exploratory operations, every pixel is characterized by its degrees of adherence to a set of classes. Therefore, the degrees of adherence of every pixel are modified once again, while taking in consideration the degree of adherence of its neighbors.

To the (n + 1) iteration, the degrees of adherence are modified like follows:

$$u_{ij}^{(n+1)} = \frac{u_{ij}^{(n)}\left(1 + q_{ij}^{(n)}\right)}{\sum_{k=1}^{C} u_{ik}^{(n)}\left(1 + q_{ik}^{(n)}\right)} \qquad (15)$$

where u_{ij} is the degree of adherence of the i point to the class j.

The correlation factor is given by :

$$q_{ij}^{(n)} = \frac{1}{cad}\sum_{v \in voisinage\ de\ i}\ \sum_{k=1}^{C} f(c, c') u_{vk}^{(n)}$$

where Card is the cardinal of the neighborhood and f is the coefficient of compatibility between the c class and c' the class. In our case we take.

$$f(c, c') = \delta_{cc'}$$

The process is repeated iteratively until its convergence.

RESULTS

We applied first the FCM algorithm on simulated data (30000 points) generated differently. Then, the modified algorithm is initialized by the result gotten following the classical FCM algorithm. For a number of classes fixed to the number of maximum classes for the classical FCM, the number of descended classes at the end of the FCM modified is very close to the real number of classes. This algorithm is especially very efficient in the press where the shapes of the classes are not regular and where the classic methods as ISODATA or K-NN leaning each on a Euclidian metric fail m ost of the time[1,7]. Indeed, in the Fig. 1, we show three densities of data overlapped (Fig. 1a), the classification by the modified FCM algorithm is shown by the Fig. 1d. Initially we have 63 prototypes represented in the Fig. 2, after a second regulation, the class's number fell again to 3 only and it corresponds to the real number of classes.

As in many problems of images segmentation, the validation of the gotten results is a delicate problem and several approaches can be considered in the goal to

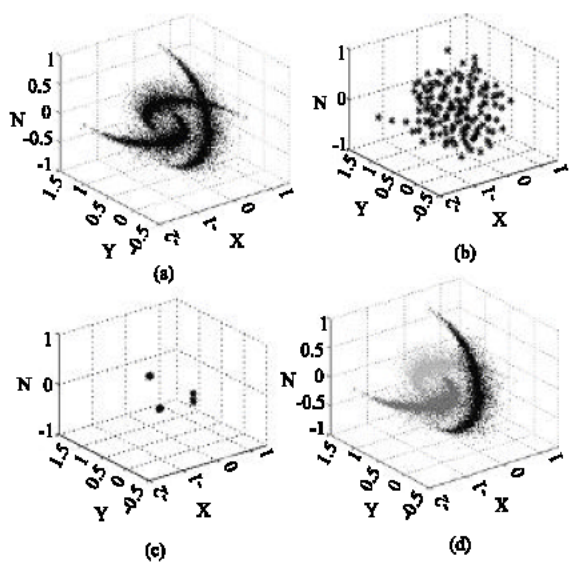

Fig. 1: (a) Set of simulated data (30000 points), (b) initial FCM clusters definition, (d) classification by modified FCM (MFCM)

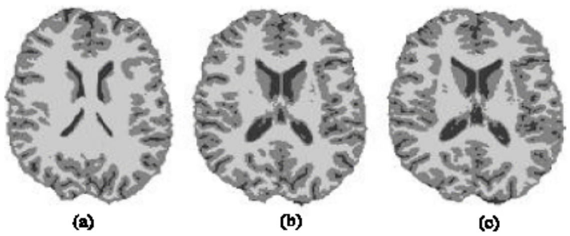

Fig. 2: (a) Classification by classical FCM, (b) classification by modified FCM, (c) classification by modified FCM followed by a fuzzy relaxation with spatial neighboring in relation to the de talairach's atlas

provide a quantitative assessment of the quality of the results. We will distinguish the approaches exploiting the result of segmentations achieved by hand by experts on real images and the approaches exploiting images of synthesis for which the reference is perfectly known mainly. In the context of the cerebral RMN images, the recourse to experts to segment real images by hand raises numerous difficulties. Indeed, the notion of localization of a contour is notably imprecise and sometimes subjective, dependent of the expert's experience. Therefore, it is very likely that the cards of references thus produced are in disagreement in numerous zones and that reliability and the reproducibility of the results are difficult to value. Besides, the complexity of the contours to detect contributes to return such an extremely trying and expensive task in time.

For the set of these reasons, the recourse to real images seems difficult to put in work in a goal of validation and it seems preferable to have resort to images of syntheses, ideally the most realistic possible.

In this work, we used the available images on the Internet site of the Montreal Neurological Institute. These images have been constructed from real pictures and are therefore realistic in terms of geometry of the contours. The intensities of tissues have been simulated with the help of physical models of the magnetic resonance process. In order to construct a realistic phantom, the effects of partial volumes have been taken in account. Thus, for every voxel of the volume, a vector describes the proportion of each of tissues that constitutes it.

The result of the classification is finally constituted by 125 volumes describing each one of the sought-after tissues (gray and white matters, cerebro-spinal liquid, grease, muscles, skull, air, etc.). Within every volume, the intensity of a voxel represents its fraction for the corresponding tissue. The use of synthesis's images to validate our approach permits to compare the segmentation gotten to a perfectly known reference card

Table 1: Rate of different tissues recognition in relation to talairach's atlas

	FCM	MFCM	MFCM + FR
WM	0.5	0.77	0.89
GM	0.53	0.69	0.78
CSL	0.64	0.76	0.79
Others	0.44	0.57	0.69

FCM: Fuzzy C-Means, MFCM: Modified Fuzzy C-Means, FR: Fuzzy Relaxation

(Atlas of Tlairach). We could also notice that an initialization by consistent FCM of an optimization by FCM modified seems very efficient for the recognition of the three cloths (WM, GM, CSL). The regulation by the fuzzy relaxation increases the rate of recognition of the different present cloths in the RMN images (Table 1).

CONCLUSION

We presented in this article a method of automatic classification of the cloths of the encephalon. The objective was to achieve this segmentation while insisting on the automatic character with the possible information minimum and while putting the accent on the quality of resulted them for a possible clinical use of our algorithm.

Our approach rests on the evaluation of the probability density by a non parametric method. A first classification is operated by the FCM algorithm, resulted them from this classification depend of a good initialization of the prototypes. To end to remedy this problem of initialization and to remain in the automatic context, we developed another FCM version whose objective function includes a second term of regulation permitting to minimize the entropy.

To increase the efficiency of our method of classification, a second correction intervenes on the picture herself, by another probabilistic regulation procedure, it is about the fuzzy relaxation that takes in consideration, this time, the spatial relation between the different voxels of our picture while using the probabilistic card constructed from a referential anatomical (here one took the referential of Talairach).

The results are especially very conclusive for the borders enter the different cloths where the algorithms of automatic classification as ISODATA or k-NN fail, our algorithm gives good resulted.

REFERENCES

1. Bensaid, A.M., L.O. Hall, L.P. Clarke and R.P. Velthuizen, 1991. MRI segmentation using supervised and unsupervised methods, Proceedings of the Thirteenth IEEE Engineering in Medicine and Biology Soc. Conference, New York: IEEE, pp: 483-489.

2. Liew, A.W.C and Hong Yan, 2003. An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation, IEEE Transactions on Medical Imaging, pp: 1063-1075.
3. Masulli, F. and A. Schenone, 1999. A fuzzy clustering based segmentation system as support to diagnosis in medical imaginh, Artificial Intelligence in Medicine, pp: 129-147.
4. Masulli, F., P. Bogus, A. Schenone and M. Artuso, 1998. Application of MEP-based fuzzy clustering to the segmentation of multivariate medical images, Proceedings of the Intl. Joint Workshop on Current Issues on Fuzzy Technologies: Methods and Environments for Planning and Programming. Singapore: World Scientific, pp: 266-277.
5. Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press.
6. Bezdek, J.C., L.O. Hall and L.P. Clarke, 1993. Review of MR image segmentation techniques using pattern recognition, Med. Phys., pp: 33-48.
7. Moussaoui, A. and V. Chen, 2004. Fuzzy automatic classification without a prior knowledge-mean shift application to MRN brain images segmentation, In symposium of electronics and telecommunications, Timisoara, Romania, pp: 157-163.
8. Schenone, A., F. Masulli and M. Artuso, 1997. A Neural Bootstrap for the Possibilistic c-means Algorithm, In: Morabito FC, Editor. Advances in Intelligent Systems. Amsterdam: IOS Press, pp: 59-66.
9. Hummel, R.A. and S.W. Zucker, 1983. On the foundations of relaxation labeling processes, IEEE Trans. PAMI, pp: 267-287.
10. Duda, R.O. and P.E. Hart, 1973. Pattern Classification and Scene Analysis, New York: Wiley.
11. Comaniciu, D. and P. Meer, 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis, IEEE Trans. Pattern Analysis and Machine Vision, Mai, pp: 1-16.
12. Parzen, E., 1962. On estimation of a probability density function and mode, Annals of Mathematic and Statistics, pp: 1065-1076.
13. Pal, N.R., K. Pal and J.C. Bezdek, 1997. A mixed c-means clustering model, In: Proceedings of the International Conference on fuzzy Systems IEEE, FUZZ-IEEE '97. Piscataway, NJ: IEEE, pp: 11-20.
14. Frigui, H. and R. Krishnapuram, 1996. A robust clustering algorithm based on competitive agglomeration and soft rejection of outliers, In CVPR, San. Fransisco, juin, pp: 550-555.
15. Beni, G. and X. Liu, 1994. A least biased fuzzy clustering method, IEEE Trans Pattern Anal Mach Intl., pp: 54-60.
16. Rose, K., E. Gurewitz and G. Fox, 1993. Constrained clustering as an optimization method, IEEE Trans Pattern Anal. Mach. Int.l, pp: 85-94.
17. Palubinskas, G. *et al.*, 1998. An unsupervised clustering method using the entropy minimization, In ICPR, Autralie, Août.