ELSEVIER

Fuzzy Sets and Systems 128 (2002) 285–303

FUZZY sets and systems

www.elsevier.com/locate/fss

Extreme physical information and objective function in fuzzy clustering

Michel Ménard *, Michel Eboueya

Laboratoire d'Informatique et d'Imagerie Industrielle, Université de La Rochelle, Avenue Marillac, 17042 La Rochelle, Cedex 1, France

Received 23 May 2000; received in revised form 4 December 2000; accepted 12 January 2001

Abstract

Fuzzy clustering algorithms have been widely studied and applied in a variety of areas. They become the major techniques in cluster analysis. In this paper, we focus on objective function models whose aim is to assign the data to clusters so that a given objective function is optimized. We propose a new approach in fuzzy clustering and show how it can be used to obtain a systematic method deriving objective functions. This approach is based on a unifying principle of physics, that of extreme physical information (EPI) defined by Frieden (Physics from Fisher Information. A Unification, 1999). The information in question is the trace of the Fisher information matrix for the estimation procedure; this information is shown to be a physical measure of disorder. We use the EPI approach for finding the effective and minimal constraint terms in objective functions. With the proposed approach we justify the constraint terms defined a priori in the Fuzzy c-means (FcM) algorithm and Possibilistic and Maximum Entropy Inference approaches. Indeed, these algorithms, by contrast, offer no such systematic method of finding its constraints. Moreover, in this context, the EPI approach derives the "reason" for the extremization of objective functions. The resulting formulae have a clearer physical meaning than those obtained by means of classical algorithms. The updated equations of our algorithm are identical to those of the possibilistic, MEI and FcM with regularization approaches.

Keywords: Measures of information; Fuzzy clustering; Classification; Extreme physical information

1. Introduction

The aim of clustering is to reduce the amount of data by grouping similar patterns together. Such grouping is pervasive in information processing. One of the motivating reasons for using clustering algorithms is to provide automated tools to help in constructing partitions into categories or taxonomies. Partitional clustering attempts to subdivide the data set into a set of subsets or clusters, which are pairwise disjoint, all non-empty, and produce the original data set via union. Objects belonging to the same cluster share common properties that distinguish them from objects belonging to other clusters. In literature, most of the clustering algorithms can be classified into two types:

- hard or crisp. In this case, the algorithm assigns each feature vector to a single cluster and ignores the possibility that this vector may also belong to

* Corresponding author. Tel.: +33-5-4645-8764; fax: +33-5-4645-8242.

E-mail addresses: michel.menard@univ-lr.fr (M. Ménard), michel.eboueya@univ-lr.fr (M. Eboueya).

0165-0114/02/$ - see front matter
PII: S0165-0114(01)00071-9

other clusters. Such algorithms are exclusive and the cluster labels are hard and mutually exclusive;
- fuzzy. Fuzzy clustering algorithms consider each cluster as a fuzzy set; while a membership function measures the degree of belonging of each feature vector in a cluster. So, each vector may be assigned to multiple clusters with some degree of sharing measured by the membership function.

Fuzzy clustering has been shown to be advantageous over crisp clustering in that total commitment of a vector to a given class is not required in each iteration. Thus, these methods are less prone to local minima than crisp clustering algorithms.

In this paper we consider the second case, more particularly the fuzzy clustering based on an objective function [14,7,27]. The objective function assigns a quality or error to each cluster arrangement, based on the distance between the data and the typical representatives of the clusters. These methods can essentially be categorized in the second level of data analysis defined in [14] with the aim to group the data on the basis of a purely qualitative investigation. The problem of fuzzy clustering is stated as follows:
- let a set of objects $Y=(y_k)_{k\in[1,n]}$ be given as the data, where $y_k=(y_{k1},y_{k2},\ldots,y_{kp})^t$ is a pattern described by p features or measures (i.e., $y_k\in\mathscr{R}^p$);
- let $\Omega=(\omega_i)_{i\in[1,c]}$ be a family of clusters (i.e., c represents the number of clusters).

Fuzzy clustering of the objects can be represented by a real $c\times n$ matrix $U=[u_{ik}]$ by interpreting u_{ik} as the degree to which y_k belongs to cluster ω_i. Let us denote $V=[v_1,\ldots,v_c]$ the matrix of cluster centers.

This paper is organized as follows. In the following subsections, we describe briefly the objective functions and constraints used in the principal fuzzy clustering algorithms. The concept of Fisher information and the principle of *extreme physical information* (EPI) are presented in Section 2. Following this context, we analyze the fuzzy clustering scenario in Section 3. A functional defined on the class of generalized characteristic functions based on the Fisher information is introduced in Section 3.1. It uses no constraint defined a priori and makes it possible to obtain an objective function connected with the situation described in fuzzy clustering problems. This is specified in Section 3.2. The general solution is obtained in Section 3.3 by resolving a second-order differential equation. The EPI approach justification for fuzzy clustering is given in Section 3.4. The results are interpreted in Section 3.5. In Section 3.6, we give a new algorithm using a 1D complex potential defined according to the observations. It derives from results obtained in the previous section. Experiments show the interests of this algorithm for which it is not necessary to define the number of clusters. Finally, Section 5 contains our conclusion.

1.1. Objective functions and constraints

The principal problem is how to formulate an objective function for the clustering problem. The objective function maps a solution to a real number measuring the quality of the solution in terms of efficiency or cost. The formulation determines how various constraints are encoded into the function. Because the optimal solution is the optimum of the objective function, the formulation defines the optimal solution. In this sense, there is an optimization problem to be solved. Among the existing clustering methods, the Fuzzy c-means [4,3] (FcM)[1] and Possibilistic c-means algorithms [18] (PcM)[2] are two of the most active and often-used data analysis methods in recent years. Both FcM and PcM algorithms described here attempt to find good cluster structure descriptors (U^*,V^*) as minimal solutions of a particular member of family of objective functions [13]:

$$J_m(U,V;Y)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m d_{ik}^2 + \sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m,$$

- where $U\in M_{\mathrm{fcn}}$ or M_{pcn}, depending on the chosen approach; these sets of all $c\times n$ non-degenerate

[1] The utilization of the FcM algorithm for various applications is well described and analyzed by Bezdek [4] and Bensaid and co-workers [2].

[2] Davé and Krishnapuram [7,17] have established an unified view of robust clustering methods and pointed out the similarities between FcM and Possibilistic approaches.

constrained matrices are defined as

$$M_{\text{fcn}} = \left\{ U \in \mathscr{R}^{c\times n} \;\middle|\; \sum_{i=1}^{c} u_{ik} = 1 \ \forall k, \quad 0 < \sum_{k=1}^{n} u_{ik} < n \ \forall i;\ u_{ik} \in [0,1] \ \forall i,k \right\},$$

where u_{ik} denotes the degree of membership of the kth pattern to the ith fuzzy cluster relative to all other clusters.

$$M_{\text{pcn}} = \left\{ U \in \mathscr{R}^{c\times n} \;\middle|\; 0 < \sum_{k=1}^{n} u_{ik} < n \ \forall i, \quad u_{ik} > 0 \ \exists i;\ u_{ik} \in [0,1] \ \forall i,j \right\},$$

where u_{ik} is the degree of representativity or typicality of the datum y_k for the cluster ω_i;

- where d_{ik} is a measure of the distance from y_k to the ith cluster prototype $v_i \in V$;
- $m > 1$ is a fuzzifier exponent which controls the extent of membership sharing between fuzzy clusters;
- $\eta = \{\eta_1, \eta_2, \ldots, \eta_c\}$, where η_i specifies the ith cluster penality term (for the FcM algorithm, $\eta_i = 0$ for $1 \leqslant i \leqslant c$). The first term in J_m is a squared error criterion. In this paper, we will consider this term as a constraint term.

Then the role of this family of objective functions permit the quantitative measure of the global quality of a solution and lead the search for a minimal solution. The clustering problem is formulated as a constrained minimization problem, whose solution depends on constraint functions that satisfy certain conditions imposed a priori:

- Fuzzy c-means algorithm uses a second constraint term, originally due to Ruspini [25], on the membership functions ($\forall y_k \in Y$, $\sum_{i=1}^{c} u_{ik} = 1$ i.e. sum to 1 over each column of U). This constraint is necessary to generate the membership update equations for an iterative algorithm: a minimization of the FcM objective function without using this constraint for the membership values results in a null solution. In the constraint form, FcM generates a fuzzy partition providing a measure of relative membership degree of each pattern to a given cluster rather than an individual degree of belonging. However, the memberships resulting from FcM do not always correspond to the intuitive concept of degree of belonging. Then, the cluster centers in fuzzy clustering or estimates for the parameters in switching regression models using an extended FcM approach (FcRM) [12] are poor. This can be a serious problem in situations where one wishes to generate membership functions from training data. However, several extended Fuzzy c-means algorithm, Fc + 2M [21] and Fc + 2RM [20], provide answers to these problems by introducing a discounting process between the classical FcM or FcRM membership functions.
- In order to overcome this problem, Krishnapuram and Keller [18] proposed the PcM algorithm, where this normalization constraint is removed. They added a supplementary term

$$\sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m \tag{1}$$

to the FcM objective function to urge a non-trivial solution for the membership values. In this case, the value u_{ik} should be interpreted as the typicality of y_k relative to cluster ω_i (i.e. a possibility of belonging). However, this constraint forces the membership values to be close to one as possible during the minimization process, thus providing a uniform unity bias [1] placed on all membership functions. Recognizing the existence of this uniform bias present in the PcM algorithm, we prefer using an alternative second term defined by Davé and Krishnapuram [7] to prevent the trivial solution as follows:

$$\sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (u_{ik} \log(u_{ik}) - u_{ik}) \tag{2}$$

with $m = 1$ for the first term in J_m. Note that $u_{ik}\log(u_{ik}) - u_{ik}$ is a monotonically decreasing function in $[0,1]$, similar to $(1 - u_{ik})^m$. This formulation of the Possibilistic approach may also be more appropriate when clusters are expected to be close to one another [18]. Using this constraint, the objective function becomes, if we substitute in J_m, constraint (1) by constraint (2):

$$J_{\text{pcm}}(U,V;Y) = \sum_{i=1}^{c}\sum_{k=1}^{n} (u_{ik}) d_{ik}^2 + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (u_{ik}\log(u_{ik}) - u_{ik}). \tag{3}$$

- Now, let us reconsider the first term in the FcM or PcM objective function:

$$\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m d_{ik}^2,$$

where weighting exponent m controls the extent of membership sharing between fuzzy clusters ($m=1$ for the chosen Possibilistic approach). This "strange" parameter, introduced by Bezdek, is unnatural and has not a physical meaning [19]. In the above objective functions, we must eliminate m, but in this case, we cannot generate the membership update equations. So Li and Mukaidono [19] proposed a new approach to fuzzy clustering by means of a *Maximum Entropy Inference* (MEI) method. Formally, this problem is written as

$$\text{maximize} \quad -\left(\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}\log(u_{ik})\right) \tag{4}$$

subject to the constraints
1. $L=\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}d_{ik}^2$.
2. $\sum_{i=1}^{c} u_{ik}=1$.

The term at extremized in (4) defines the *entropy* of a probability density function. The maximum entropy criterion is generally used when we have only prior information. This concept will be precised in Section 3.5.3. The resulting formulae have a more beautiful form and a clearer physical meaning than those obtained by means of the Fuzzy c-means method. However, the probabilistic constraint is also necessary to generate the membership update equations. Thus, in this case it is also necessary defining the constraint terms a priori.

1.2. Aims of this paper

Whether in determining cluster or in estimating cluster centers, the existing techniques are essentially based upon the a priori choice of the formulation (i.e., constraints and information terms) of the objective function. In this paper, we explicitly give a criterion (obtained from three axioms) for providing theoretical justification of objective functions, constraints, membership functions and potential functions in fuzzy clustering. As described previously, in classical algorithms of extremization, there are some constraints that are imposed to satisfy certain conditions defined a priori. The FcM algorithm, the Possibilistic and Maximum Entropy Inference approaches offer no systematic method of finding their constraints. Moreover, the objective functions have no significance in its own right.

We observe as intrinsic data the vectors $y_i=v_i+x_i$, $i=1,\ldots,c$ (i.e., the data fluctuations $(x_1,\ldots,x_c)$ are presumed to characterize solely the phenomenon under measurement, x_i is a p-vector). Any fluctuation $y_i-v_i=x_i$ should occur with a probability that is independent of the absolute size of v_i:

$$p_i(y_i/v_i)=p_i(x_i), \quad x_i=y_i-v_i.$$

Our aim is to propose a physical approach to finding:
- the effective and minimal constraint terms included in the objective function adapted to the fuzzy clustering problem. J represents these constraint terms;
- the objective function K;
- the distribution $p_i(x)$ relative to the given measurement scenario (i.e., membership functions).

Our work is inspired by the work of Frieden in which he defined and developed a unifying principle of physics, that of *extreme physical information*. EPI provides a mechanism to finding the constraint terms and seeking an exact solution for the unknown distribution of the measurement scenario, whose data have Fisher information level I. I arises out of a "flow" $J\rightarrow I$ of information from the measured phenomenon to the data space, where J embodying all unknown constraints that are imposed by the physical phenomenon under measurement (i.e., stored within the system). I and J are functionals which depend upon the values of the unknown distribution $p_i(x)$. This relay of information from the phenomenon to the data defining an information loss $K\equiv I-J$ (i.e., physical information of the system). $K\equiv I-J=\text{extrem}$ constitutes a variational principle for finding J and $p_i(x)$. Aside its advantage in providing a mechanism for finding constraint terms, EPI derives the "reason" for the extremization of K.

For our purposes, it is useful and easier to work in the continuous representation space (i.e., with an infinite number of elements) with an integral rather than in the discrete domain (i.e., finite number of objects) with a sum. The integral, K (i.e., the objective

function in discrete domain), has the form

$$K[\mathbf{q}(\mathbf{x})] \equiv I[\mathbf{q}'] - J[\mathbf{q}, \mathbf{x}], \quad \mathbf{q}, \mathbf{x} \text{ real},$$

$$x_i = (x_{i1}, \ldots, x_{iv}, \ldots, x_{ip}), \quad \mathrm{d}\mathbf{x} \equiv \mathrm{d}x_1 \ldots \mathrm{d}x_p,$$

$$\mathbf{q} \equiv (q_1, \ldots, q_i, \ldots, q_c),$$

$$\mathbf{q}'(\mathbf{x}) \equiv \frac{\partial q_1}{\partial x_{i1}}, \frac{\partial q_1}{\partial x_{i2}}, \ldots, \frac{\partial q_c}{\partial x_{cp}},$$

where $I[\mathbf{q}']$ is the trace of the Fisher information matrix obtained from data for the system. Its formulation is known; $J[\mathbf{q}, \mathbf{x}]$ represents all constraints to be determined; $p_i(x) \equiv q_i(x)^2$. The functions q_i are probability amplitudes, i.e., whose squares are unknown probability densities. The probability amplitude q_i verifies the following elementary properties:

Property 1

$$\lim_{|x_i| \to \infty} \frac{\mathrm{d}q_i}{\mathrm{d}x_i} q_i = 0. \tag{5}$$

Property 2

$$\lim_{|x_i| \to \infty} q_i = 0. \tag{6}$$

In order to get rid of heavy notations, we will name as $K[\mathbf{q}] \equiv K[\mathbf{q}(x)]$, $q_i \equiv q_i(x_i)$ the integral and the distributions we want to estimate. K is a functional. It depends upon the values of one or more functions q_i continuously over the domain of $\mathbf{x}$, their derivatives with respect to all the x_{iv}, and $\mathbf{x}$. Then our approach based upon an information criterion allows us to obtain constraint terms J, the unknown distribution $p_i(x)$, $i = 1, \ldots, c$, and the objective function defined by the *physical information* K adapted to fuzzy clustering problem. This approach leads to membership functions or potential functions.

2. Extreme physical information and Fisher information

The unifying principle of physics, that of EPI, defined by Frieden [10] permits physics to be viewed within a unified framework of measurement. Some implications of this approach in the analysis of complex systems have been considered by Frieden. Indeed the EPI theory was introduced to provide a scheme for handling a variety of physics problems in which a fundamental role is played by the observer into the phenomenon that he observes: statistical mechanics and thermodynamics, quantum mechanics, the Einstein field equations, and quantum gravity...

Consider the problem of estimating c vectors $v_i = (v_{i1}, v_{i2}, \ldots, v_{ip})$, $i = 1, \ldots, c$. Any fluctuation $y_i - v_i = x_i$ should occur with a probability

$$p_i(y_i/v_i) = p_i(x_i), \quad x_i = y_i - v_i.$$

Frieden shows that the accuracy in the estimates of the c parameters v_i is determined by the Fisher information I that has some useful physical properties [9] (i.e., Fisher information may be regarded as a physical measure of disorder which is related to experimentation). It provides new definitions of disorder, time and temperature, and a variational approach finding a multiple-component probability density function law $p_i(x_i)$ of a vector variable x_i.

The Fisher information[3] I in a multi-parameter, multi-component measurement scenario obeys [10]:

$$I[\mathbf{p}] = \sum_i \int \mathrm{d}x_i \frac{1}{p_i(x_i)} \sum_v \left(\frac{\partial p_i(x_i)}{\partial x_{iv}} \right)^2. \tag{7}$$

This expression further simplifies if we introduce real probability amplitude $q_i(x_i)$ [10]:

$$I[\mathbf{q}] = 4 \sum_i \int \mathrm{d}x_i \sum_v \left(\frac{\partial q_i}{\partial x_{iv}} \right)^2, \tag{8}$$

where $q_i \equiv q_i(x_i)$ is the ith component probability amplitude for the fluctuation $x_i = (x_{i1}, \ldots, x_{ip})$ in the measurement. They are the basic unknown items of the problem; $q_i^2(x_i) \equiv p_i(x_i)$ denotes the probability density function for the noise value x_i.

$I[\mathbf{q}]$ is called the *intrinsic* information, since it is a functional of the probability amplitudes that are intrinsic to the measured phenomenon. Under certain conditions, information I obeys an *I-theorem*:

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} \leqslant 0 \tag{9}$$

[3] This information is not the Fisher information per se, but rather the trace of the Fisher information matrix for the system. It is actually an upper bound to the Fisher information. Frieden calls it *the information capacity* for the estimation procedure.

with t the time. So that, at equilibrium, $I = \min$. This means that I is a monotonic measure of system disorder. As for the Boltzmann entropy the equilibrium solution must be the stationary solution that attains in I the smallest minimum among all possible minima. We use this theorem in Section 3.

Frieden defines a second information item called *bound Fisher information* $J[\mathbf{q}]$ to embody all constraints that are imposed by the physical phenomenon under measurement. $J[\mathbf{q}]$ and $\mathbf{q}$ are generally found by applying the principle of *Extreme Physical Information* consisting of three axioms [10]: *conservation law*, *existence of information densities*, *zero-condition on the microscopic level*. These axioms imply the variational principle and the zero-condition comprising the overall principle:

- the variational principle is the extremization of the functional for finding $\mathbf{q}$

$$K[\mathbf{q}] = I[\mathbf{q}] - J[\mathbf{q}] = \text{extrem}. \tag{10}$$

  There is no loss of the perturbed Fisher information in its relay from the phenomenon to the intrinsic data (i.e., due to a measurement, the system is perturbed, causing a perturbation δJ in the bound information, so it must be that $\delta I = \delta J$, necessarily $\delta(I - J) = 0$);
- the zero-condition

$$I[\mathbf{q}] - \kappa J[\mathbf{q}] = 0, \quad \kappa \leqslant 1 \tag{11}$$

  is the zeroing of the functional. It is possible to find with the previous relation the unknown information functional $J[\mathbf{q}]$. Information $J[\mathbf{q}]$ is ultimately identified by an invariance principle that characterizes the measured phenomenon. In Section 3, for defining objective functions, we will use a Fourier transformation as an invariance principle connecting input space x with another, conjugate space, x'. The existence of such a unitary transformation guarantees the validity of the EPI method for fuzzy clustering [10].

In the following section, we analyze the problem of finding the formulation of the optimal objective functions (i.e., the optimal membership functions) under an optimality criterion defined from EPI scheme. To attain this aim, we give the following definitions.

Definition 1 (Kreinovich et al. [16]). A final criterion is defined as a criterion that chooses a unique optimal family of functions q_i (i.e., a family that is better with respect to the criterion than any other family).

Definition 2. The criterion chosen to determine a family of optimal objective functions in fuzzy clustering is defined by the Frieden's scheme: application of (1) variational principle, (2) zero-condition and (3) I-theorem.

If this criterion is final in the sense of Definition 1, it means that this criterion must build up an objective function leading to the search for an optimal solution. Proving that is one of the aims of Section 3.

3. Extreme physical information in fuzzy clustering

3.1. Scenario and EPI scheme

We first analyze the following scenario. Denote the true positions of the centroid vectors as v_i, $i = 1, \ldots, c$, where $v_i = (v_{i1}, v_{i2}, \ldots, v_{ip})$ are the usual Cartesian components. A fuzzy clustering problem has the added difficulty that y is unlabeled (i.e., for a given data y, the cluster to which y belongs is not known). In this case, the observation y (y real, $y = (y_1, \ldots, y_p)$) obeys

$$y = v_i + r_i, \quad i = 1, \ldots, c, \tag{12}$$

where r_i, a p-vector, is the residual associated with observation y and v_i (i.e. fluctuations). Notice we use the notation r_i instead of x_i here.

The structure of the EPI problem is one of finding a probability assignment, the corresponding probability amplitude $q_i(r_i)$ (denote $q_i^2(r_i) \equiv u_i(y)$) which avoids bias, while agreeing with whatever information is given, $u_i(y)$ lies in $[0, 1]$ and that verifies the principle of extreme physical information defined previously. [4] For this scenario, by Eq. (8), the Fisher information quantity is

$$\mathscr{I}[\mathbf{q}] = \sum_{i=1}^{c} 4 \int \mathrm{d}r_i \sum_{v} \left(\frac{\partial q_i}{\partial r_{iv}} \right)^2. \tag{13}$$

[4] First, the centroid vectors, v_i were assumed to be constant vectors. So the basic unknowns of the problem are the real amplitude functions $q_i(r_i)$, $i = 1, \ldots, c$.

Since r_i's do not range over the same values, we can apply the EPI scheme on each independent scenario $r_i = y - v_i,\ i = 1,\ldots,c$ (i.e., c independent informations). Then we have

$$\mathscr{I}[\mathbf{q}] = \sum_{i=1}^{c} I[q_i], \tag{14}$$

where

$$I[q_i] = 4\int \mathrm{d}r_i \sum_{v=1}^{p} \left(\frac{\partial q_i}{\partial r_{iv}}\right)^2. \tag{15}$$

Corresponding to each information functional $I[q_i]$ is a bound information functional $J[q_i]$ (to be found) which verifies the EPI approach for the c scenario:

- variational principle implies

$$I[q_i] - J[q_i] = \text{extrem}, \tag{16}$$

- zero-condition gives

$$I[q_i] - \kappa J[q_i] = 0. \tag{17}$$

The functions q_i of $r_i,\ i = 1,\ldots,c$ (probability amplitudes, i.e., whose squares are also membership functions) are to be found.

If we denote the bound information functional $J[q_i]$ as follows:

$$J[q_i] = 4\int \mathrm{d}r_i\, j_i(q_i, r_i),$$

where j_i may be regarded as an information density. Then, by Eq. (15), Extremum principle, given by Eq. (16), becomes for q_i:

$$I[q_i] - J[q_i] = 4\int \mathrm{d}r_i \left[\sum_{v=1}^{p}\left(\frac{\partial q_i}{\partial r_{iv}}\right)^2 - j_i(q_i, r_i)\right] = \text{extrem}. \tag{18}$$

In our case, we consider a general form for j_i (i.e., depending on q_i and r_i). The zeroing of Eq. (17) (*Assumption* 1: we used $\kappa = 1$ for a suitable solution of the problem, see Section 3.4 for a justification of this assumption) gives

$$I[q_i] - J[q_i] = 4\int \mathrm{d}r_i \left[\sum_{v=1}^{p}\left(\frac{\partial q_i}{\partial r_{iv}}\right)^2 - j_i(q_i, r_i)\right] = 0. \tag{19}$$

Integrating by parts $I[q_i]$ and using Property 1, the zero-condition $I[q_i] - J[q_i]$ becomes

$$-4\int \mathrm{d}r_i \left[q_i \sum_{v=1}^{p} \frac{\partial^2 q_i}{\partial r_{iv}^2} + j_i(q_i, r_i)\right] = 0. \tag{20}$$

The answer to these variational problems, given by Eqs. (18) and (20), is that $q_i(r_i)$ must obey Euler–Lagrange equations. Then, we obtain the following solutions (for $i = 1,\ldots,c$):

$$\sum_{v=1}^{p} \frac{\partial^2 q_i}{\partial r_{iv}^2} = -\frac{1}{2}\frac{\partial j_i(q_i, r_i)}{\partial q_i} \tag{21}$$

and

$$q_i \sum_{v=1}^{p} \frac{\partial^2 q_i}{\partial r_{iv}^2} = -j_i(q_i, r_i). \tag{22}$$

By using Eqs. (21) and (22), the information density solution $j_i(q_i, r_i)$, which satisfies the variational principle (18) and the zero-condition (20), is obtained by integration:

$$j_i(q_i, r_i) = q_i^2(r_i) f_i(r_i)$$

for some functions $f_i(r_i)$. Substituting the previous result into Eq. (21) or (22) produces the same solution of the form of

$$\nabla^2 q_i(r_i) = -q_i(r_i) f_i(r_i), \tag{23}$$

where the Laplacien ∇^2 is with respect to coordinates of r_i. The form of $f_i(r_i)$ directly affects that of $q_i(r_i)$ (and $u_i(y)$).

We assume that

$$f_i(r_{iv}) = \alpha_{iv} - \beta_i r_{iv}^2, \quad \alpha_{iv}, \beta_i \text{ constant } \forall v = 1,\ldots,p$$

for the following reasons:

Assumption 2: it is reasonable to require that the membership function must not change if we apply an arbitrary rotation around v_i in p-dimensional space $\mathscr{R}^p$ (i.e., the criterion is rotation-invariant). So, the membership function should not depend upon the sign of $r_i = y - v_i$. Hence $u_i(y)$ should be even in each component r_{iv};

Assumption 3: Since we do not have any reason to believe that some measurements are more accurate than others, similarly it is reasonable to assume that $u_i(y)$ should depend upon each r_{iv} in the same way

(i.e., covariance requires that the system functionally depends upon all its coordinates—including the time, if that is one of the coordinates—in the same way). Therefore, α_{iv}'s are the same for all the r_{iv}: $\alpha_{iv} = \alpha'_i$, $v = 1, \ldots, p$;

Assumption 4: it is also reasonable to assume that the measurement errors are relatively small (i.e., r_i is the residual associated with observation y and v_i). So that we can neglect the terms that are beyond the quadratic.

These considerations imply that $f_i(r_i)$ should be expandable as a power series in even powers of r_i,

$$f_i(r_i) = \alpha_i - \beta_i r_i^2, \tag{24}$$

where α_i is proportional to α'_i. The constants α_i and β_i are to be defined. At this point, conditions 1–4 are merely assumptions. They lead to useful properties for the objective functions in fuzzy clustering: centroids are *unbiased*, coincident solutions for maximum entropy inference and extreme physical information approaches (cf. Section 3.3). However, we will justify below these assumptions in Section 3.4 by using a Fourier transformation as an invariance principle.

3.2. Constraint determination

Using this quadratic form in Eq. (18) leads to an extremization problem:

$$\begin{aligned} &I[q_i] - J[q_i] \\ &\quad = 4 \int \mathrm{d}r_i \left[\sum_{v=1}^{p} \left(\frac{\partial q_i}{\partial r_{iv}} \right)^2 - q_i^2(\alpha_i - \beta_i r_i^2) \right] \\ &\quad = \text{extrem}, \end{aligned} \tag{25}$$

Eq. (25) is equivalent to, $i = 1, \ldots, c$ ($\beta_i = \text{const}$):

$$\begin{aligned} &\frac{1}{4\beta_i}(I[q_i] - J[q_i]) \\ &\quad = \int \mathrm{d}r_i \left[\sum_{v=1}^{p} \frac{1}{\beta_i} \left(\frac{\partial q_i}{\partial r_{iv}} \right)^2 - q_i^2 \left(\frac{\alpha_i}{\beta_i} - r_i^2 \right) \right] \\ &\quad = \text{extrem}. \end{aligned}$$

Since the r_i's do not range over the same values, they are independent. So the EPI problem leads to find $\{q_i\}_{i \in [1,c]}$ obeying:

$$K'[\mathbf{q}] = \sum_{i=1}^{c} \frac{1}{4\beta_i}(I[q_i] - J[q_i]) = \text{extrem}. \tag{26}$$

For our purposes, it is useful to work with a discrete form (with a finite number of elements):

$$\begin{aligned} K'[U] = &\sum_{i=1}^{c} \frac{1}{4\beta_i} I_i^f(u_i) \\ &- \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \left[\frac{\alpha_i}{\beta_i} - d^2(y_k, v_i) \right] = \text{extrem}, \end{aligned}$$

where we defined $u_{ik} \equiv u_i(y_k) = q_i^2(r_{ik})$, $d^2(y_k, v_i) = r_{ik}^2$. $I_i^f(u_i)$ and u_{ik} are, respectively, the discrete form of Fisher information $I[q_i]$ and the probability density $q_i^2(r_i)$. Next, consider the second term in the previous equation:

$$\begin{aligned} J^f(U, V; Y) &= \sum_{i=1}^{c} \frac{1}{4\beta_i} J(u_i, v_i; Y) \\ &= -\sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \left[\frac{\alpha_i}{\beta_i} - d^2(y_k, v_i) \right] \end{aligned}$$

represents all constraints that are imposed by the scenario. On the other hand, this term is

$$\begin{aligned} J^f(U, V; Y) = &-\sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} \frac{\alpha_i}{\beta_i} \\ &+ \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} d^2(y_k, v_i). \end{aligned}$$

If we denote $\eta_i = \alpha_i / \beta_i$, this relationship is

$$\begin{aligned} J^f(U, V; Y) = &-\sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} u_{ik} \\ &+ \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} d^2(y_k, v_i). \end{aligned}$$

At this point, we must point out that the constraints are those of the objective function defined a priori by Krishnapuram and Keller [18] in the particular implementation of the possibilistic approach described

in Introduction. We also remark that the EPI approach leads to an objective function defined by a first information term $\sum_{i=1}^{c} 1/4\beta_i I_i^f[u_i]$ intrinsic to the measured phenomenon and another information term embodying all constraints that are imposed by the physical phenomenon under measurement.

By summary, let j_{epi} be the information density found by applying the Frieden's scheme defined in Section 2. EPI method leads to an objective function representing the following c-independent objective functions:

$$\frac{1}{4\beta_i} I_i^f[u_i] - \sum_{k=1}^{n} j_{\text{epi}}(u_{ik}, d_{ik}; \eta_i) = \text{extrem}, \quad i = 1, \ldots, c, \tag{27}$$

with $j_{\text{epi}} = u_{ik} g_i(d_{ik})$, $g(d_{ik}) = \eta_i - d_{ik}^2$ and $\eta_i, \beta_i =$ const. The first term defines the discrete form of the Fisher information. This term gives a global measure of the physical information of the system. The meaning of Eq. (27) is different from the one of classical methods because no constraint a priori is needed in order to define it.

The possibilistic approach, defined by Krishnapuram, implies an objective function representing the following c-independent objective functions:

$$\eta_i I_i^s[u_i] - \sum_{k=1}^{n} j_{\text{pcm}}(u_{ik}, d_{ik}; \eta_i) = \min, \quad i = 1, \ldots, c, \tag{28}$$

with $j_{\text{pcm}} = j_{\text{epi}}$. The first term defines the entropy of Shannon,

$$I_i^s[u_i] = \sum_{k=1}^{n} u_{ik} \log u_{ik} \tag{29}$$

whose range is the set of nonnegative real numbers.

In comparing the entropy method with the EPI approach, it is to be noted that I_i^f is a local measure of u_{ik} and I_i^s is a global measure of this one.

3.3. General solution

This calculus is quite similar to Frieden [10] in the determination of the Maxwell–Boltzmann velocity law, more particularly momentum probability amplitudes.

Substituting Eq. (24) into Eq. (23) gives

$$\nabla^2 q_i(r_i) + q_i(r_i)[\alpha_i - \beta_i r_i^2] = 0. \tag{30}$$

That is, the answer q_i is the solution to a second-order differential equation. We use a method of separation of variables for solving this partial differential equation by making a substitution of the form

$$q_i(r_i) = q_{i1}(r_{i1}) q_{i2}(r_{i2}) \ldots q_{ip}(r_{ip}) \tag{31}$$

and breaking the resulting equation into a set of p independent ordinary differential equations:

$$\frac{\mathrm{d}^2 q_{iv}(r_{iv})}{\mathrm{d}r_{iv}^2} + (\alpha_i' - \beta_i r_{iv}^2) q_{iv}(r_{iv}) = 0,$$

where $v = 1, \ldots, p$, $\sum_{v=1}^{p} \alpha_i' = p\alpha_i' = \alpha_i$ (Assumption 2).

Now letting

$$\alpha_i' = k_i^2 \quad \text{and} \quad \beta_i = \lambda_i^2$$

gives

$$\frac{\mathrm{d}^2 q_{iv}(r_{iv})}{\mathrm{d}r_{iv}^2} + (k_i^2 - \lambda_i^2 r_{iv}^2) q_{iv}(r_{iv}) = 0. \tag{32}$$

Eq. (32) is called Weber differential equation. Functions

$$q_{iv}(r_{iv}) = \exp\left\{ -\left(\frac{\lambda_i}{2} r_{iv}^2 \right) \right\} 2^{-n/2p} H_n(\sqrt{\lambda_i} r_{iv}) \tag{33}$$

are solutions of Eq. (32) if the constants verify:

$$k_i^2 = 2\lambda_i (n + \tfrac{1}{2})$$

or

$$\alpha_i' = 2\sqrt{\beta_i}(n + \tfrac{1}{2}),$$

H_n are the Hermite polynomials, (n is a positive integer).

Eqs. (31) and (33) give the solutions

$$q_i^n(r_i) = A_i \exp\left\{ -|r_i|^2 \left(\frac{\lambda_i}{2} \right) \right\} 2^{-n/2} (H_n(\sqrt{\lambda_i} r_{iv}))^p, \tag{34}$$

where A_i is a constant.

As described by Frieden [10], the solution which attains the absolute minimum in the Fisher information

$I[q_i]$ is the solution at equilibrium (i.e., the Gaussian solution in our case). Indeed the presence of Hermite polynomials in the solution (34) causes subsidiary maxima and minima (H_n admits $n+1$ extrema). Since $I[q_i]$ contains any derivatives (cf. Eq. (15)), extrema cause an increased gradient. Following this remark, the solution at equilibrium of Eq. (30) is this one giving the smallest minimum $I[q_i]$ among all possible minima (i.e., given by the Minimum Fisher information, and obtained by setting $n=0$ in Eq. (34). This implies $\alpha_i' = \sqrt{\beta_i}$).

Then, the general solution is a Gaussian solution:

$$q_i(r_i) = A_i \exp\left\{-|r_i|^2\left(\frac{\lambda_i}{2}\right)\right\}, \tag{35}$$

A_i is a constant.

Thus we have the following result for each independent scenario $y = v_i + r_i$, $i = 1,\dots,c$ and with the α_i's, β_i's positive reals:

Let the following final criterion 1 be:
Assumptions 2–4 implying

$$j_i(q_i, r_i) = q_i^2(\alpha_i - \beta_i r_i^2).$$

Variational principle implying

$$4\int \mathrm{d}r_i \left[\sum_{v=1}^{p}\left(\frac{\partial q_i}{\partial r_{iv}}\right)^2 - j_i(q_i, r_i)\right] = \text{extrem.}$$

Zero condition and Assumption 1 implying

$$-4\int \mathrm{d}r_i \left[q_i \sum_{v=1}^{p}\frac{\partial^2 q_i}{\partial r_{iv}^2} + j_i(q_i, r_i)\right] = 0.$$

Minimum Fisher information implying

$$4\int \mathrm{d}r_i \sum_{v=1}^{p}\left(\frac{\partial q_i}{\partial r_{iv}}\right)^2 = \min.$$

Theorem 1. *If $r_i \in \mathscr{R}^p$, and $\alpha_i = p\sqrt{\beta_i}$ for $i \in [1,c]$, $\forall A_i \in \mathscr{R}$ the probability amplitudes $q_i(r_i)$, $i = 1,\dots,c$, may verify the final criterion defined previously only if*:

$$q_i(r_i) = A_i \exp\left\{-|r_i|^2\frac{\lambda_i}{2}\right\}, \quad i = 1,\dots,c \tag{36}$$

and λ_i checks: $\lambda_i = \alpha_i/p$, $i = 1,\dots,c$.

In a discrete form, we obtain $\forall i \in [1,c]$:

$$q_i(r_{ik}) = A_i \exp\left\{-\frac{d^2(y_k, v_i)}{2a_i^2}\right\}, \tag{37}$$

with $a_i^2 = 1/\lambda_i$.

Using $u_{ik} = q_i^2(r_{ik})$, we now express u_{ik}, $i = 1,\dots,c$ as follows:

$$u_{ik} = C_i \exp\left\{-\frac{d^2(y_k, v_i)}{\eta_i}\right\}, \tag{38}$$

where C_i depends on A_i, and $\eta_i = a_i^2 = \text{const}$.

Comments

- With $C_i = 1$, Eq. (38) produces the same solution than the possibilistic approach;
- In Section 3.4, we will give a justification of the EPI approach where the Assumptions 1–4 will not be useful to define the information density j_{epi}. This justification leads to a second more general theorem where the requirement of j_{epi} (i.e., assumptions) is not necessary;
- Thus, our general approach provides a precise physical justification for the constrain terms used in the Possibilistic and MEI approaches;
- q_i, $i = 1,\dots,c$ are independent probability amplitude;
- In the case of the normalization of u_{ik}, wished a posteriori, and by setting $C_i = A$ and $\eta_i = 2\sigma^2$ $\forall i \in [1,c]$, then the probability densities verify

$$\sum_{i=1}^{c} A \exp\left(-\frac{d^2(y_k, v_i)}{2\sigma^2}\right) = 1.$$

It is elementary that this equation implies

$$A = \frac{1}{\sum_{i=1}^{c}\exp(-d^2(y_k, v_i)/2\sigma^2)}. \tag{39}$$

Substituting result (39) into Eq. (38) produces the same solution than the MEI approach:

$$u_{ik} = \frac{\exp(-d^2(y_k, v_i)/2\sigma^2)}{\sum_{j=1}^{c}\exp(-d^2(y_k, v_j)/2\sigma^2)}. \tag{40}$$

3.4. EPI approach justification for fuzzy clustering

Here, a one-dimensional analysis is given for simplicity. Extensions to p-dimensional scenario are

straightforward. Following the scenario described in Section 3.1, the observation y obeys

$$y = v_i + r_i, \quad i = 1, \ldots, c,$$

where v_i, $i = 1, \ldots, c$ represent the true positions of the centroid vectors and r_i is the residual, (i.e., a random excursion) associated with observation y and v_i, whose probability amplitude law $q_i(r_i)$ is sought. The measurement perturbs the amplitudes $q_i(r_i)$, $i = 1, \ldots, c$ of the problem and starts EPI going as a process. The information expression becomes in one dimension and for the cluster ω_i (r_i's do not range over the same values, so we can apply the EPI scheme on each independent scenario $r_i = y - v_i$, $i = 1, \ldots, c$):

$$I[q_i] = 4 \int \mathrm{d}r_i \left(\frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right)^2, \quad i = 1, \ldots, c. \tag{41}$$

We use a Fourier transform as invariance principle in order to guarantee the validity of the EPI approach [10] for fuzzy clustering.

Let $\Phi_i(\mu_i)$ be the characteristic function defined as the inverse Fourier transform (denoted $\mathscr{F}^{-1}$) of the probability density function $p_i(r_i)$. Let $\phi_i(\mu_i)$ be the function such as

$$\phi_i(\mu_i) * \phi_i(\mu_i) = \Phi_i(\mu_i), \tag{42}$$

where $*$ denotes the convolution product. ϕ_i also uniquely specifies the probability density function $p_i = q_i^2$. One can consider this function as a function characterizing q_i. Using this definition, ϕ_i verifies

$$\phi_i(\mu_i) = \mathscr{F}^{-1}[q_i(r_i)].$$

We define a Fourier transform space consisting of functions $\phi_i(\mu)$ of μ_i obeying

$$q_i(r_i) = \frac{1}{2\pi} \int \mathrm{d}\mu_i \phi_i(\mu_i) \mathrm{e}^{-\mathrm{j}\mu_i r_i}. \tag{43}$$

The unitary nature of this transformation guarantees the validity of the EPI variational procedure. One of the interests of the Fourier transform is that it analyzes the global regularity of q_i: the smoother a function (i.e., the lower the number of continuous derivatives), the more compact its Fourier transform. Naturally, this is a global regularity condition. By differentiating (43) and applying Parceval's theorem, we obtain

$$\int \mathrm{d}r_i \left| \frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right|^2 = \int \mathrm{d}\mu_i |\phi_i(\mu_i)|^2 \mu_i^2, \quad i = 1, \ldots, c. \tag{44}$$

Using (44) in (41) gives

$$I[q_i] = 4 \int \mathrm{d}\mu_i |\phi_i(\mu_i)|^2 \mu_i^2, \quad i = 1, \ldots, c. \tag{45}$$

When the invariance principle is the statement of a unitary transformation between measurement space and conjugate coordinate space, then the solution to requirement *zero-condition*, given by Eq. (17), is that functional J be simply the re-expression of I in the conjugate space. Thus we have:

$$I = 4 \int \mathrm{d}\mu_i |\phi_i(\mu_i)|^2 \mu_i^2 = J, \quad i = 1, \ldots, c. \tag{46}$$

This is the invariance principle for the given scenario. The same value of I can be expressed in the new space (μ) where it is called J. J is then the bound information for the scenario. With this invariance principle, we justify Assumption 1: $\kappa = 1$ (cf. Eq. (19)).

A single observation y yields the result that r_i is to be found in the distance interval according to v_i: $r_i \ldots r_i + \mathrm{d}r_i$. The corresponding probability is $|q_i(r_i)|^2 \, \mathrm{d}r_i$. The uncertainty (i.e., the mean value of the 2nd power[5] of r_i by considering that the corresponding probability is $|q_i|^2$) ΔR and the conjugate uncertainty $\Delta \Xi$ are chosen as

$$(\Delta R)^2 = \int_{\Re} r_i^2 |q_i|^2 \, \mathrm{d}r_i, \tag{47}$$

$$(\Delta \Xi)^2 = \int_{\Re} \mu_i^2 |\phi_i|^2 \, \mathrm{d}\mu_i. \tag{48}$$

Proposition 1. $\forall q_i \in L^2$, $\exists (\alpha_i, \beta_i) \in \mathscr{R}^{2*+}$,

$$\beta_i (\Delta R)^2 + (\Delta \Xi)^2 = \alpha_i \int |q_i(r_i)|^2 \, \mathrm{d}r_i. \tag{49}$$

Proof. Since q_i is a finite energy function (property required for a probability amplitude), $(\Delta R)^2$ and $(\Delta \Xi)^2$ exist. One only needs, for example, that

$$\beta_i = 1 \tag{50}$$

[5] (ΔR) and $(\Delta \Xi)$ include most of the total energy but do not contain all of it.

and

$$\alpha_i = \frac{(\Delta R)^2 + (\Delta \Xi)^2}{\int |q_i(r_i)|^2 \, \mathrm{d}r_i}. \tag{51}$$

End of proof. Then, by Eq. (49), the conjugate uncertainty $\Delta\Xi$ obeys

$$(\Delta \Xi)^2 = \alpha_i \int |q_i(r_i)|^2 \, \mathrm{d}r_i - \beta_i (\Delta R)^2. \tag{52}$$

Using Eqs. (52), (46) and (41), Extremum principle becomes for q_i:

$$I[q_i] - J[q_i] = 4 \int \mathrm{d}r_i \left(\frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right)^2 - \left[\alpha_i \int |q_i(r_i)|^2 \, \mathrm{d}r_i - \beta_i (\Delta R)^2 \right] = \text{extrem}.$$

Hence, we have shown (cf. Eq. (25)) that, for $i = 1, \ldots, c$:

$$I[q_i] - J[q_i] = 4 \int \mathrm{d}r_i \left[\left(\frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right)^2 - (\alpha_i - \beta_i r_i^2) q_i^2(r_i) \right]. \tag{53}$$

So that $j_i(q_i, r_i) = q_i^2(\alpha_i - \beta_i r_i^2)$ (cf. Section 3.2). Thus we have the following result for each independent scenario $y = v_i + r_i$, $i = 1, \ldots, c$:

Let the following final criterion 2 be:
Variational principle implying

$$4 \int \mathrm{d}r_i \left[\sum_{v=1}^{p} \left(\frac{\partial q_i}{\partial r_{iv}} \right)^2 - j_i(q_i, r_i) \right] = \text{extrem}.$$

Zero condition implying

$$-4 \int \mathrm{d}r_i \left[q_i \sum_{v=1}^{p} \frac{\partial^2 q_i}{\partial r_{iv}^2} + j_i(q_i, r_i) \right] = 0.$$

Minimum Fisher information implying

$$4 \int \mathrm{d}r_i \sum_{v=1}^{p} \left(\frac{\partial q_i}{\partial r_{iv}} \right)^2 = \min.$$

Theorem 2. *If* $r_i \in \mathscr{R}^p$ *for* $i \in [1, c]$, $\forall A_i \in \mathscr{R}$ *the probability amplitudes* $q_i(r_i)$, $i = 1, \ldots, c$, *may verify the final criterion defined previously only if*:

$$q_i(r_i) = A_i \exp\left\{ -|r_i|^2 \frac{\lambda_i}{2} \right\}, \quad i = 1, \ldots, c,$$

$$\lambda_i = \text{const}, \quad i = 1, \ldots, c.$$

3.5. EPI interpretation for fuzzy clustering

3.5.1. Coincident global minimum

In summary, the membership functions, u_i, $i = 1, \ldots, c$ obtained by EPI, Possibilistic and MEI approaches are similar. In Possibilistic and MEI approaches the formulation of the objective function is given with a priori constraints. Moreover, the normalization constraint is necessary in MEI approach to obtain membership functions. In EPI approach this constraint is not necessary. Then it can be used a posteriori. As we saw in Section 3.2, EPI approach leads to the following c-independent objective functions:

$$\frac{1}{4\beta_i} I_i^f[u_i] - \sum_{k=1}^{n} j_{\text{epi}}(u_{ik}, d_{ik}; \eta_i) = \text{extrem}, \quad i = 1, \ldots, c. \tag{54}$$

Possibilistic approach implies an objective function representing the following c-independent objective functions:

$$\eta_i I_i^s[u_i] - \sum_{k=1}^{n} j_{\text{pcm}}(u_{ik}, d_{ik}; \eta_i) = \min, \quad i = 1, \ldots, c, \tag{55}$$

with $j_{\text{pcm}} = j_{\text{epi}}$. It is easy to show that, for the general solution (37), Fisher information and Shannon information verify the relation (as $\beta_i = 1/\eta_i^2$):

$$\frac{1}{4\beta_i} I_i^f[q_i = A \mathrm{e}^{-r_i^2/2\eta_i}] = -\eta_i I_i^s[p_i = A \mathrm{e}^{-r_i^2/\eta_i}]. \tag{56}$$

Then, the general solution gives a coincident global minimum[6] for the EPI objective function and Possibilistic approach using two different information terms: I_i^f and I_i^s (cf. Fig. (1)).

[6] By using Legendre's condition, one can show that the solution of EPI objective function Eq. (54) is also a minimum.

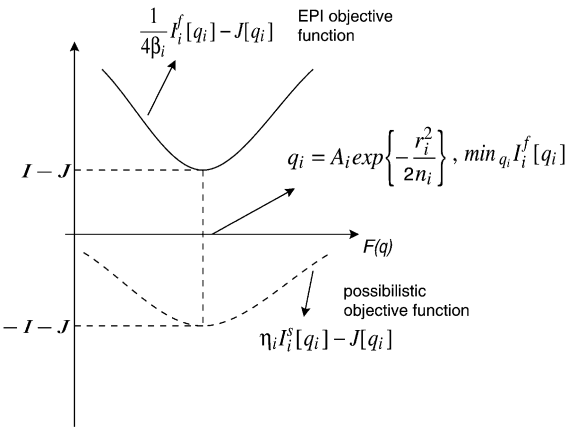

Fig. 1. The general solution gives a coincident global minimum for the EPI objective function and Possibilistic approach using two different information terms: I_i^f and I_i^s. $F(q)$ denotes the set of families of probability amplitudes.

If the prototype of each cluster can be represented by the cluster center v_i, then it can be easily verified that the prototype update equation for the objective function, defined by the approach EPI, can be written as

$$v_i = \frac{\sum_{k=1}^{n} u_{ik} y_k}{\sum_{k=1}^{n} u_{ik}}. \tag{57}$$

Eqs. (38) and (57) can be used in an alternative fashion in an iterative algorithm to estimate the cluster centers v_i.

3.5.2. Underlying constraint and uncertainties

For simplicity, we use one-dimensional notation. The approach is easily generalized to p dimensions. Let $(\Delta R)_N$ and $(\Delta \Xi)_N$ be the normalized uncertainties such as

$$(\Delta R)_N^2 = \frac{(\Delta R)^2}{\int |q_i|^2 \, \mathrm{d}r_i}, \tag{58}$$

$$(\Delta \Xi)_N^2 = \frac{(\Delta \Xi)^2}{\int |\phi_i|^2 \, \mathrm{d}\mu_i}. \tag{59}$$

Proposition 2. *$\exists(\alpha_i, \beta_i) = (\zeta, \zeta^2)$, $\zeta \in \mathscr{R}^{*+}$, such as $\forall q_i \in L^2$, if q_i verifies the final criterion 2, then the following relation between uncertainties is also verified: $(\Delta R)_N(\Delta \Xi)_N = \frac{1}{2}$ (i.e., a balance is attained between the uncertainties $(\Delta R)_N$ and $(\Delta \Xi)_N$).*

Proof. If q_i verifies the final criterion 2, then by Eqs. (19), it exists $\alpha_i = \alpha_i' = \sqrt{\beta_i}$, such that q_i verifies (zero condition in one-dimensional case):

$$I[q_i] - J[q_i] = 4 \int \mathrm{d}r_i \left[\left(\frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right)^2 - \alpha_i q_i^2 + \beta_i r_i^2 q_i^2 \right] = 0. \tag{60}$$

By setting $\zeta = \alpha_i$, and by Eq. (60), q_i verifies

$$K[q_i, \zeta] = I[q_i] - J[q_i, \zeta] = \int \left| \frac{\mathrm{d}q_i}{\mathrm{d}r_i} \right|^2 \mathrm{d}r_i - \zeta \int |q_i|^2 \, \mathrm{d}r_i + \zeta^2 \int r_i^2 |q_i|^2 \, \mathrm{d}r_i = 0,$$

where ζ is a still-to-be-determined real number. With definitions (47) and (48), we may use Parceval's theorem to get

$$\int |\phi_i|^2 \, \mathrm{d}\mu_i = \int |q_i|^2 \, \mathrm{d}r_i. \tag{61}$$

With these definitions, $K[q_i, \zeta]$ can be simplified to yield:

$$K_N[q_i, \zeta] = (\Delta \Xi)_N^2 + \zeta^2 (\Delta R)_N^2 - \zeta = 0, \tag{62}$$

where $K_N[q_i, \zeta]$ represents normalized $K[q_i, \zeta]$. Since this equation is verified by q_i obeying the final criterion 2, we can now determine the value of ζ for which K_N is minimum. If we differentiate Eq. (62) with respect to ζ, we obtain the condition for a minimum:

$$2\zeta_{\min}(\Delta R)_N^2 - 1 = 0,$$

from which we find immediately

$$\zeta_{\min} = \frac{1}{2(\Delta R)_N^2}.$$

So, we obtain

$$(\Delta R)_N^2 (\Delta \Xi)_N^2 = \tfrac{1}{4} \tag{63}$$

or

$$\Delta R_N \Delta \Xi_N = \tfrac{1}{2}.$$

Comment: for $\alpha_i = \zeta_{\min}$, Proposition 1 is also verified. □

3.5.3. Fuzzy clustering with regularization

In this subsection, we discuss fuzzy clustering problem from the viewpoint of regularization of an ill-posed problem. A problem is said to be ill-posed if it fails to satisfy one or more of the following criteria: solution exists, is unique and depends continuously on the data. Additional prior assumptions have to be imposed on the solution to convert an ill-posed problem into a well-posed one. Recently, new methods of fuzzy c-means have been proposed using this concept. In analogy with the maximum entropy principle, these approaches give preference to a solution which maximizes a given measure of entropy or equivalently, minimizes the corresponding information measure. Maximum entropy solution is then equivalent to minimizing information subject to constraints given by the objective function of FcM clustering. In [23], Miyamoto and Mukaidono show that the concept of regularization plays an important role in the FcM. Indeed, they considered Fuzzy c-means to be a regularization for the Crisp c-means. They regard extremal points $\{0,1\}$ as singular solutions, whereas intermediate points are considered to be a regular solution[7] (i.e., m is a smooth factor). In [22], they have reformulated the MEI approach by considering the entropy term as a regularization term. Introducing this term, they show they minimize a functional (given by FcM objective function) depending on the desired solution regularized by minimizing the total amount of information. In this context, they consider that PcM and possibilistic approaches may be viewed as another regularization. In [15], Ichihashi et al. show that EM algorithm for gaussian mixture model can be derived from FcM clustering by introducing regularization with K–L information instead of entropy. We show in [21] the concept of regularization extended to fuzzy clustering of overlapping clusters.

In comparing these entropy-term-based methods with the EPI approach, it is to be noted that the Fisher information is a regularization local measure of u_{ik} and the entropy of shannon is a global measure of this one. Indeed, EPI approach is based on an information function that controls derivative values. It is clear that minimizing the averaged Fisher information will have the effect of "smoothing" the data. Besides its advantage in providing a mechanism for the constraint term determination, EPI derives the "reason" for the extremization of the objective function. It does not have to assume maximum ignorance of some type as in Maximum entropy approaches, for example. At this point, we must point out that classical statistical systems, where the probability laws are simple exponential functions, a local regularization is equivalent to a global regularization (i.e., Maximum entropy and EPI approaches agree in this case).

3.5.4. Harmonic potential in fuzzy clustering

It is interesting to write Eq. (30), found in Section 3.3, under a more general form by using the framework of quantum mechanics, and more particularly that of the Schrödinger's equation. For the fuzzy clustering scenario and by final criterion 2, the general solution q_i is the solution to a second-order differential equation.

$$\nabla^2 q_i(r_i) + q_i(r_i)[\alpha_i - \beta_i r_i^2] = 0. \tag{64}$$

By setting, $\alpha_i = k_i^2$, $\beta_i = \lambda_i^2$, $V(r_i) = \lambda_i^2 r_i^2$ and $E = k_i^2$ the general solution q_i obeys

$$\nabla^2 q_i(r_i) + q_i(r_i)[k_i^2 - \lambda_i^2 r_i^2] = 0 \tag{65}$$

or

$$\nabla^2 q_i(r_i) + q_i(r_i)[E - V_i(r_i)] = 0. \tag{66}$$

The Schrödinger's stationary (i.e., time-independent) equation, in the presence of a potential field, can be written as:

$$\frac{\hbar^2}{2m}\nabla^2\psi(r) + \psi(r)[E - V(r)] = 0, \tag{67}$$

where the wave function ψ determines the spatial probability $p(r) = |\psi(r)|^2$. ψ must be finite, single-valued, and continuous, its derivative with respect to position must also be finite, single-valued and continuous. Moreover ψ vanishes at

[7] Another viewpoint is considered by Rose et al. in [24]. They give a probabilistic interpretation of maximum entropy using a deterministic annealing approach to clustering. The resulting algorithm is started with high fuzziness (i.e., equal membership in all clusters) at high temperatures and converges to the non-fuzzy basic ISODATA algorithm at low temperatures).

infinity.[8] $V(r)$ is the potential energy of the particle as function of position r. The eigenvalue E corresponds to a particular allowable energy state of the particle and ψ is the corresponding eigenvector. In this case, the general solution $q_i,\ i=1,\ldots,c$, given by the Schrödinger's equation, is the wave function ψ_0 corresponding to the lowest energy E_0:

$$\psi_0(r) = C\exp(-\tfrac{1}{2}\lambda r^2), \tag{68}$$

where the constant C can be determined by the normalization condition. The associated eigenvalue is

$$E_0 = \frac{\lambda}{k^2}. \tag{69}$$

It is also interesting to note that the final criterion 2 yields to the Schrödinger's stationary equation in the presence of c harmonic potential fields $V_i(r_i) = \lambda_i^2 r_i^2$ (α_i^2 is the corresponding eigenvalue) i.e., each cluster center generates a harmonic potential $V_i(r_i)$. The encoding of the smoothness prior in terms of derivatives including in Fisher information leads to isotropic potential functions.

3.6. Schrödinger's time-independent equation

3.6.1. Potential function defined from observations

In the previous subsection, we saw that the fuzzy clustering process required harmonic potentials. More complex potential functions also appear to be immediately useful. For example, they are used in the Mountain method [26] and in various improved implementations of this method [6,5]. In this section, we give a new algorithm using a 1D complex potential defined according to the observations $y_k,\ k=1,\ldots,n$. Experiments show the interests of this algorithm for which there is no need to know the number of clusters.

Let y be an unspecified point for which one wants to calculate the potential. To calculate the potential to this point, it is necessary to evaluate an estimator of the local density around the observations $y_k,\ k=1,\ldots,n$. We use the method described in [8,11]. Let the potential function $V(y)$ be given for an unspecified y point (cf. Appendix A). $V(y)$ is characterized by its local properties. This potential function assumes that physical properties in a neighborhood of space present some coherence and generally do not change abruptly. For a very dense population, the potential in boundary of this area increases very quickly. For a very wide population the potential increases slowly.

We first present two simple examples to provide insights into generated potential functions. Here we discuss the shape of potential function $V(y)$ in the 1-D case. Fig. 2a shows the potential function obtained from observations y_k. These observations are represented by symbol '+', and are generated from a gaussian cluster. Fig. 2b illustrates the behavior of the potential function $V(y)$ found according to observations generated from two gaussian clusters.

Now the remaining problem is to determine the probability amplitudes q_i (or ψ_i), $i=1,\ldots,c$.

3.6.2. Wave functions and fuzzy clustering

Here, we proceed with the solution of the Schrödinger's time-independent equation in the presence of the potential function $V(y)$ defined previously. As we saw in the previous section, the Schrödinger's time-independent equation, in the one dimension case and in the presence of a potential field, is

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}y^2} + V\psi = E\psi\,, \tag{70}$$

where V and ψ are function of y. $V(y)$ is the potential energy, $\psi(y)$ the wave function, and E energy associated with the wave function. m is the mass of the particle and $\hbar$ the Planck's constant.

Since the Schrödinger's equation includes several constants, it is customary to write $\hbar^2/2m = 1$. Then, Eq. (70) becomes:

$$-\frac{\mathrm{d}^2\psi}{\mathrm{d}y^2} + V\psi = E\psi\,. \tag{71}$$

In the case of a complex potential, $V(y)$, Eq. (71) can only be solved approximately (cf. Appendix).

The following experiments illustrate the procedure for finding the probability density function $p_i \equiv |\psi_i|^2$ from two examplary instances and demonstrate

[8] So that the wave function can be normalized, i.e., $\int |\psi|^2\,\mathrm{d}r = 1$.

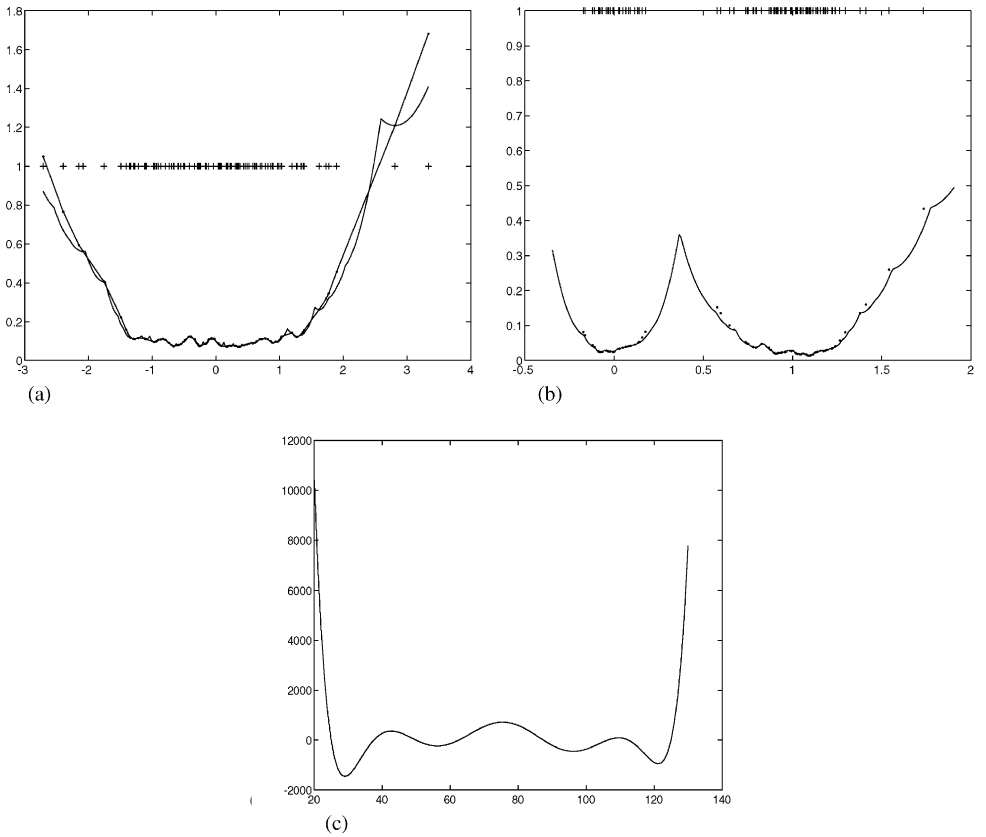

Fig. 2. Examples showing potential functions. The potential function translates the densities of observations into variable size areas. (a) This example illustrates a situation containing one cluster. The potential function is defined for all points y and is obtained from observations y_k generated from a gaussian cluster. Observations are represented by symbol '+'. (b) This example illustrates a situation containing two clusters. (c) A typical example of potential function is shown. Potential function is of the form: $V(y)=\prod_{i=1}^{8}(y-\gamma_i)$.

the use of potential functions in our fuzzy clustering approach. The experiments are conducted as three parts:

- given a set of objects or observations $Y=(y_k)_{k\in[1,n]}$;
- computing the potential function $V(y)$ according to these observations by using Eqs. (73) and (74);
- proceeding to the solution of the Schrödinger's time-independent equation in the presence of the potential field $V(y)$ in order to determine the probability density function $|\psi_i(y)|^2$ (or $|q_i(y)|^2$).

In Fig. 3, we present two examples that provide a clear computation process.

- A typical example of potential function is depicted in Fig. 2c. The considered potential function is of the form: $V(y)=\prod_{i=1}^{8}(y-\gamma_i)$. The probability density functions (i.e., $|\psi_i(y)|^2$), solutions of the Schrödinger's equation are plotted in Fig. 3a. Those which present a well located state are presented (their shape do not present oscillations). In this example, we obtain four probability density functions characterizing four clusters. At each solution, is associated an energy.
- In Fig. 2b, the example deals with a situation in which there are two clusters: it can be seen that the potential function $V(y)$ translates the densities of observations into variable size areas. Step 2: proceeding to the solution of the Schrödinger's equation in the presence of this potential gives two probability density functions (cf. Fig. 3b) (these functions present a well located state).

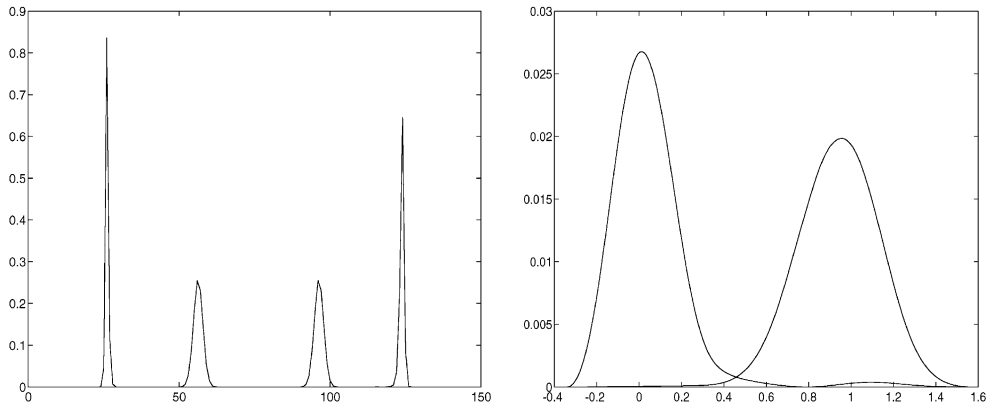

Fig. 3. The probability density functions, solutions of the Schrödinger's equation are plotted. Those which present a well localized state are presented (their shapes do not present oscillations). At each solution, is associated an energy. (a) In this example, we obtain four probability density functions characterizing four clusters. The associated potential field is described in Fig. 2c. (b) We obtain two probability density functions characterizing two clusters. The associated potential field is described in Fig. 2b.

The previous results seem to indicate that the performance of this algorithm is superior (i.e., it is not necessary to define the number of clusters). However it is to be borne in mind that this algorithm is computationally and implementationally more complex. Indeed it is prohibitory for applications where a pattern is described by $p>3$ measures.

4. Conclusion

In fuzzy clustering, the principal problem is how to formulate an objective function. The formulation determines how various constraints are encoded into the function. In this article, we give a final criterion leading to a family of optimal objective functions (for this criterion). This one is based on the Fisher information, I-theorem and the EPI principle of Frieden: *variational principle* and *zero-condition*. We justify the application of this principle to the problem of fuzzy clustering. This approach offers a systematic method allowing to obtain and to justify the constraints of the possibilistic, MEI and FcM approaches. We remark that the EPI approach leads to an objective function defined by a first information term intrinsic to the measured phenomenon and another information term embodying all constraints that are imposed by the fuzzy clustering scenario. The update equations are identical to that of the possibilistic, MEI or FcM with regularization (in the case of the wished normalization a posteriori). It is also shown that uncertainties on the observations and defined on the conjugate space check a balance. The product of these uncertainties is minimum. The solution obtained by our approach is similar to the solution of the Schrödinger's stationary equation in the presence of c harmonic potential fields generated by the cluster centers. Then we give an algorithm using a 1D complex potential defined according to the observations. Even if this algorithm is computationally and implementationally more complex, it is interesting because there is no need to know the number of clusters.

Acknowledgements

The authors would like to thank the anonymous reviewers and B. Roy Frieden for the valuable comments and suggestions.

Appendix A. Potential function defined from observations and wave function

Let y be an unspecified point for which one wants to calculate the potential. First of all, let us introduce

the vector d^l, whose kth component is the average distance from the observation y_k with its l closer neighbors y_{kr}, $r=1,\ldots,l$

$$d_k^l \equiv d^l(y_k) = \frac{1}{l}\sum_{r=1}^{l} d(y_k, y_{kr}), \tag{A.1}$$

where $1/d^l(y_k)$ is an estimator of the local density around the point y_k [8]. l plays a significant role:

- a great value of l, implies a smoothing and decreasing the local properties;
- a low value introduces noise into the values $d^l(y_k)$.

Let the potential function $V(y)$ be given for an unspecified y point [8,11]:

$$V(y) = \min_{k=1,n}\left[d^l(y_k) + \frac{D(y_k)}{d^l(y_k)^2} d^2(y_k, y)\right], \tag{A.2}$$

with

$$D(y_k) = \frac{1}{l}\sum_{r=1}^{l} |d^l(y_k) - d^l(y_{kr})|, \tag{A.3}$$

$V(y)$ is characterized by its local properties. In this function, two terms appear:

- $d^l(y_k)$, mean distance from y_k with its l closer neighbors,
- $d^2(y, y_k)$, the distance between y and y_k multiplied by $D(y_k)/d^l(y_k)^2$.

$D(y_k)$ controls the sharpness of the potential function [8,11]. It represents the variation of local density in the area centered on the observation y_k. The denominator $d^l(y_k)^2$ is a term of modulation on $D(y_k)$: increases in the effect of $D(y_k)$ in a zone of high density and reciprocally.

Now, we proceed with the solution of the Schrödinger's time-independent equation in the presence of the potential function $V(y)$ defined previously. In the case of a complex potential, $V(y)$, Eq. (71) can only be solved approximately.

Given $V(y)$, the probability amplitudes ψ_i can be estimated by proceeding to the solution of the Schrödinger's equation. The chosen numerical analysis method consists in dividing the interval of integration into n small finite length intervals δ_y, such as: $\psi^j = \psi((j-1)\delta_y)$ and $V^j = V((j-1)\delta_y)$.

So, in the discrete case where the surface is sampled at discrete points, we use the second-order difference to approximate the second-order derivative and Eq. (71) becomes

$$-\frac{\psi^{j+1} - 2\psi^j + \psi^{j-1}}{\delta_y^2} + V^j\psi^j = E\psi^j. \tag{A.4}$$

By setting

$$H = \begin{pmatrix} \frac{2}{\delta_y^2} + V^1 & -\frac{1}{\delta_y^2} & 0 & 0 & \cdots \\ -\frac{1}{\delta_y^2} & \frac{2}{\delta_y^2} + V^2 & -\frac{1}{\delta_y^2} & 0 & \cdots \\ 0 & -\frac{1}{\delta_y^2} & \frac{2}{\delta_y^2} + V^3 & -\frac{1}{\delta_y^2} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

The 1D Schrödinger's equation forms a tri-diagonal matrix. It is an eigenvalue equation and such that the solutions to Schrödinger's equation appear as a sequence of eigenenergies and eigenfunctions:

$$\mathbf{H}\psi = E\psi. \tag{A.5}$$

Implicitly $\psi_0 = \psi_{n+1} = 0$ was fixed, which corresponds to the assumption that the wave function vanishes at infinity. Mathematically, one replaced a differential equation with boundary conditions by a problem of eigenvectors.

Comments on the physical significance of the eigenvalues of $\mathbf{H}$:

Let the operator $\mathbf{H}$ with a complete, orthonormal system of eigenfunctions ψ_m be given with eigenvalues a_m and a wave function

$$\psi(x) = \sum_m c_m\psi_m(x). \tag{A.6}$$

Let the system be in the state given by (A.6). The result of a measurement can only be one of the eigenvalues a_m. The probability of measuring a_m is $|c_m|^2$. Thus, the possible values of the energy are the energy eigenvalues $E_m = a_m$.

The potential function translates the densities of observations into potential wells. The tendency to bind a pattern (i.e., particle) will certainly grow with the depth and width of a potential well. The probability density functions which present a well localized state (with the lowest energy) correspond to the ground states (one per potential well).

References

[1] A. Schneider, Weighted possibilistic c-means clustering algorithms, Fuzz'IEEE 2000, 7 May 2000, Texas, USA.

[2] A.M. Bensaid, L.O. Hall, J.C. Bezdek, L.P. Clarke, M.L. Silbiger, J.A. Arrington, R.F. Murtagh, Validity-guided (re) clustering with applications to image segmentation, IEEE Trans. Fuzzy Systems 4 (2) (1996) 112–123.

[3] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[4] J.C. Bezdek, S.K. Pal, Fuzzy Models for Pattern Recognition, IEEE Press, Piscataway, NY, 1992.

[5] S. Chiu, Selecting input variables for fuzzy models, J. Intell. Fuzzy Systems 4 (1996) 243–256.

[6] S. Chiu, J.J. Cheng, Fuzzy model identification based on cluster estimation, J. Intell. Fuzzy Systems 2 (3) (1994) 267–268.

[7] R.N. Davé, R. Krishnapuram, Robust clustering methods: a unified view, IEEE Trans. Fuzzy Systems 5 (2) (1997) 270–292.

[8] B. Dubuisson, Diagnostic et Reconnaissance des Formes, Paris hermes edition, 1990.

[9] B.R. Frieden, Fisher information as a measure of time, Astrophysics and Space Sciences, Vol. 244, Kluwer Academic Publishers, Dordrecht, 1996, pp. 387–391.

[10] B. Roy Frieden, Physics from Fisher Information. A Unification, Cambridge University Press, Cambridge, 1999.

[11] D. Grenier, Méthode de détection d'évolution. Application a l'instrumentation nucléaire, These de Docteur Ingenieur, Université de Compiegne, CNRS Heudiasyc, 1984.

[12] R.J. Hathaway, J.C. Bezdek, Switching regression models and fuzzy clustering, IEEE Trans. Fuzzy Systems 1 (3) (1993) 195–203.

[13] R.J. Hathaway, J.C. Bezdek, Optimization of clustering criteria by reformulation, IEEE Trans. Fuzzy Systems 3 (2) (1995) 241–245.

[14] F. Hoppner, F. Klawonn, R. Kruse, T. Runler, Fuzzy Cluster Analysis, Wiley, New York, 1999.

[15] H. Ichihashai, K. Honda, N. Tani, Gaussian mixture pdf approximation and fuzzy c-means clustering with entropy regularization, Proc. of the 4th Asian Fuzzy System Symp., Tsukuba, Japan, 31 May–3 June 2000, pp. 217–221.

[16] V. Kreinovich, H.T. Nguyen, Y. Yam, Optimal choices of potential functions in fuzzy clustering, Technical Report CUHK-MAE-98-001, The Chinese University of Hong Kong, January 1998.

[17] R. Krishnapuram, J.M. Keller, The possibilistic c-means algorithm: insights and recommendations, IEEE Trans. Fuzzy Systems 4 (3) (1996) 148–158.

[18] R. Krishnapuram, A possibilistic approach to clustering, IEEE Trans. Fuzzy Systems 1 (2) (1993) 98–110.

[19] R. Li, M. Mukaidono, A maximum entropy approach to fuzzy clustering, Conf. on Fuzzy Systems (Fuzz'IEEE), Yokohama, Japan, 20–24 March 1995.

[20] M. Ménard, Fuzzy clustering and switching regression models using ambiguity and distance rejects, Fuzzy Sets and Systems, to be published.

[21] M. Ménard, C. Demko, P. Loonis, The fuzzy c+2-means: solving the ambiguity reject in clustering, Pattern Recognition 33–37 (2000) 1219–1237.

[22] S. Miyamoto, An overview and new methods in fuzzy clustering Conf. on Knowledge-based Intelligent Electronic Systems, Adelaide, Australia, 21–23 April 1998, pp. 33–40.

[23] S. Miyamoto, M. Mukaidono, Fuzzy c-means as a regularization and maximum entropy approach, Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA'97), Prague, Chech., Vol. 2, 25–30 June 1997, pp. 86–92.

[24] K. Rose, E. Gurewitz, G. Fox, A deterministic annealing approach to clustering, Pattern Recognition Lett. 11 (1990) 589–594.

[25] E. Ruspini, A new approach to clustering, Inform. and Control 15 (1969) 22–32.

[26] R.R. Yager, D.P. Filev, Approximate clustering via the mountain method, IEEE Trans. Systems, Man Cybernet. 24 (8) (1994) 1279–1284.

[27] M.S. Yang, A survey of fuzzy clustering, Math. Comput. Model. 18 (11) (1993) 1–16.