Chapter 12
Reinforced Operators in Fuzzy Clustering Systems

Andrei Doncescu, Sebastien Regis, and Nabil Kabbaj

Summary Knowledge-based systems need to deal with aggregation and fusion of data with uncertainty. To use many sources of information in numerical forms for the purpose of decision or conclusion, systems are supposed to have tools able to represent the knowledge in a mathematical form. One of the solutions is to use fuzzy logic operators.

12.1 Introduction

Modeling a nonlinear dynamic system can be achieved using differential equation or time series describing the behavior of the system based on the knowledge cause/effect. Although most identification methods assume that the input and output variables of the process are known, in reality it is often not clear which variables should be considered as input to the model. It seems obvious to introduce the knowledge about the system as background and to be able to deal with it in a fusion context. The most easy way is to express the knowledge in a linguistic form which mathematically expresses uncertainty. Several mathematical frameworks exist for modeling uncertainty:

1. probability theory,
2. belief theory,
3. fuzzy sets and possibility theory.

A. Doncescu (✉) · N. Kabbaj
LAAS-CNRS, University of Toulouse, 7, avenue du Colonel Roche, 31077 Toulouse, France
e-mail: adoncesc@laas.fr

N. Kabbaj
e-mail: nkabbaj@laas.fr

S. Regis
Grimaag-Guadeloupe, Campus de Fouillole, B.P. 592, 97157 Pointe Pitre Cedex, France
e-mail: sregis@univ-ag.fr

F. Xhafa et al. (eds.), *Complex Intelligent Systems and Their Applications,*
Springer Optimization and Its Applications 41,
DOI 10.1007/978-1-4419-1636-5_12, © Springer Science+Business Media, LLC 2010

The probability theory is the most commonly used. In this field, the a posteriori probabilities for an object to be a member of a class are computed according to Bayes decision rule. This probabilistic approach is criticizable for several reasons: initially the construction of a probability distribution requires much more information than an expert is able to provide. The choice of a parameterized family of distribution functions mainly results from a concern to simplify calculation. Consequently, the fit of the model to the expert opinion is debatable. An expert prefers to provide intervals rather than isolated values because his knowledge is of limited reliability and with accuracies. Belief theory handles inaccurate and uncertain information. The possibility theory and fuzzy sets are based on fuzzy logics. Fuzzy logics are characterized as "logics based on the real number." In these types of logic, one considers that the truth degrees are taken from the real line $\mathbb{R}$. Fuzzy modeling and control are typical examples of techniques that make use of human knowledge and deductive process basically using inference mechanisms. The advantage of fuzzy modeling is that the information can be either of numerical or of symbolic nature. Its representation as numerical degrees leads to a quantification of its characteristics (uncertain, imprecise, incomplete) which have to be taken into account in a fusion process. Therefore, the kernel of these mechanisms is the fusion operator defined as

$$F(x_1, x_2, x_3, \ldots, x_n), \tag{12.1}$$

where x_i denotes the representation of information issued from sensor i. The goal of the fusion operators (or aggregation) is to carry out the fusion of information resulting from various and varied sources, having the goal to take a better decision than from one source only, by reducing imprecision and uncertainty and increasing completeness. Therefore the fusion or aggregation is an reinforced information.

Information-fusion methods are executed in two ways:

1. combining each source in a parallel way: in this case, all data to be merged are available at the time of fusion, or
2. reviewing information: data are available at different moments, and the decision to be taken is updated as soon as new information comes in.

12.2 Fusion Operators

There is a great number of fusion operators. The choice and utilization of an aggregation operator depend on many parameters. This choice depends on the *fusion* itself. Before going further, it is important to give the definition of fusion suggested by Bloch and Hunter (2001):

> *The fusion consists to join together or aggregate the information coming from various sources, and to exploit this new resulting information in various applications like the decision, numerical estimation, etc.*

This definition points out two principal elements. First, the definition emphasizes the combination of the information. Then the accent is put on the fusion itself. Another

important aspect is to know what kind of data we seek to aggregate. There are two types of data: one is proposed by Bloch and Hunter (2001), and the other is presented by Dubois and Prade (2004). The latter data consist of the following:

– *The observations*. They describe the world from more or less particular point of view. We speak in general about numerical data provided by sensors.
– *The knowledge*. It describes how the world is *in general*. In this case we speak often about data from human observations rather than from sensors.
– *The preferences.* It is information which describes how we like the world to be. Of course, this information is coming from persons.
– *The regulations*. We speak about generic information presented on the form of laws.

According to the definition given in Dubois and Prade (1994) and Yager and Rybalov (1996), for any fusion operator F of two variables, one says that:

1. F is conjuctive if $F(x, y) \leq \min(x, y)$;
2. F is disjunctive if $F(x, y) \geq \max(x, y)$;
3. F behaves like a compromise if $x \leq F(x, y) \leq y$ if $x \leq y$ and $y \leq F(x, y) \leq x$ otherwise.

If all sources are reliable, it is possible to use a conjunctive fusion. But if some sources are reliable and some are not, or if their reliability is unknown, then it is better to use disjunctive fusion. A weighted logical combination can also be applied to merge data sources that have different degrees of reliability.

Bloch has been proposed the following classification to describe the operators in terms of their behavior (Bloch 1994):

1. Context-Independent Constant-Behavior (CICB) operators: they have the same behavior whatever the values of the information to combine.
2. Context-Independent Variable-Behavior (CIVB) operators: the behavior depends on the values of x and y.
3. Context-Dependent (CD) operators: they depend on a global knowledge or measure on the sources to be fusioned.

12.3 Fusion Operators in Fuzzy Sets and Possibility Theory

The operators used in these theories are CICB and are constituted by three families:

1. triangular norms T-norms $\rightarrow T$;
2. triangular conorms Γ-conorms $\rightarrow C$;
3. mean operators M.

A continuous T-norm is a continuous binary operation "$*$" on the real unit square which is:

1. commutative;
2. associative;

3. nondecreasing and having 1 for its unit element.

A continuity property is often added to these properties. T-norms generalize intersection to fuzzy sets. Examples of T-norms are:

1. minimum T-norm $x * y = \min(x, y)$, introduced by Dummett;
2. $x * y = \max(0, x + y - 1)$;
3. product T-norm $x * y = xy$.

It is easy to prove the following result: for any T-norm T, the following inequality holds:

$$\forall (x, y) \in [0, 1]^2, \quad T(x, y) \leq \min(x, y). \tag{12.2}$$

This result shows that the "min" is the greatest T-norm which has a conjunction behavior. This kind of operators is used when all the sources are reliable. We remark that in Probability and Dempster–Shafer, the operators are the product and the orthogonal sum, respectively, having a conjunction behavior and are CICB.

A T-conorm is defined as an operator $C : [0, 1]^2 \to [0, 1]$ such that C is commutative, associative, monotonic, and admitting 0 as unit element. It is easy to prove the following inequality for any T-conorm C:

$$\forall (x, y) \in [0, 1]^2, \quad C(x, y) \geq \max(x, y). \tag{12.3}$$

Disjunctive operators are used when at least one source is reliable. The other sources could be uncertain. The most known is the maximum.

A particular class is the uninorms proposed by Yager and Rybalov (1996). They are commutative, associative, and having a neutral element $e \in [0, 1]$, which the user may fix. In practice, an uninorm is often defined by a T-norm on the interval $[0, e]$ and by a T-conorm on the interval $[e, 1]$. We point out that some uninorms are symmetrical sums (Dubois and Prade 2004). It is pointed out that the symmetrical sums, introduced by Silvert (1979), are operators whose characteristic is to be symmetrical concerning a subset and its complement. Similar to the uninorms, the hybrid operators are the combinations of T-norms and T-conorms named mixed connectives. The goal of these operators is to get the advantages of T-norms and T-conorms by the variation of a parameter. These operators have been studied by Zimmerman and Zynso (1980) and later by Piera-Carreté et al. (1988). The most known are:

– the linear connective:

$$\alpha T(x_1, \ldots, x_n) + (1 - \alpha) C(x_1, \ldots, x_n); \tag{12.4}$$

– the geometrical connective:

$$T(x_1, \ldots, x_n)^{\alpha} + C(x_1, \ldots, x_n)^{(1-\alpha)}, \tag{12.5}$$

where $0 \leq \alpha \leq 1$.

The zero norms (*nullnorms*) defined by Calvo et al. (2001) are commutative, associative operators, having an absorbent element $a \in [0, 1]$, which the user can fix a priori.

The means operators provide a value between the minimum and the possible maximum. A mean operator M is defined as a function such that:

1. $\min(x, y) \leq M(x, y) \leq \max(x, y)$;
2. $M(x, y) = M(y, x)$;
3. M is increasing w.r.t. both arguments.

We notice that the aggregation operators need to satisfy the monotonic condition. We could interpret this condition by the fact that if the marginal information increases, the numerical value also increases or, in a less strong condition, does not decrease.

Let us quote, for example, the arithmetic and geometrical means or the median. Ordered weighted averages suggested by Yager (1988) (*Ordered Weighted Averaging: OWA*) are also mean operators into which it is possible to introduce a weighting depending on the importance and on the reliability of the sources. The class of operators OWA has the advantage of being stable reporting to positive linear transformation. The Min and Max are particular cases of OWA operators, but the most important property of these operators is the ability to represent optimistic or pessimistic attitude.

12.4 The Triple Π Operator

Suppose that for a given class, a vector form has membership degrees important for all features considered. In the human reasoning, an aggregation of all this marginal information will be higher than each degree taken separately (Yager and Rybalov 1998). In this type of reasoning, the membership degrees which are strong will be reinforced mutually. This behavior is called *positive reinforcement*. Of a similar reasoning, if for a given class, an object has small membership degree, the aggregation will be weaker than weakest of the membership degree values. We speak in this case about *negative reinforcement*.

The total reinforcement is a property which translates certain aspects of the human reasoning. Using the operator having this property can thus be interesting in measurement where we seek a system close to this type of reasoning. The completely reinforced operators are a particular class of operators having the characteristic to be both positively and negatively reinforced. The concept of reinforcement was presented by Yager and Rybalov (1998). The only completely reinforced operator that we know is the triple Π developed by these two authors (Yager and Rybalov 1998). We figure out that this triple Π is also a symmetrical sum.

Definition 12.1 An aggregation operator L whose arguments are within the interval $[0, 1]$ has the property of positive reinforcement if when all its attributes are affirmative (i.e., greater than or equal to 0.5), it satisfies the conditions

$$L(x_1, \ldots, x_n) \geq \max_i \big[L(x_i)\big]. \tag{12.6}$$

Similarly, an aggregation operator L whose arguments are within the interval [0, 1] has the property of negative reinforcement if when all its attributes are nonaffirmative (i.e. lower than or equal to 0.5), it satisfies

$$L(x_1, \ldots, x_n) \leq \min_i \big[L(x_i)\big]. \tag{12.7}$$

An operator having the above two properties is defined as being *totally reinforced (fully reinforced)*.

The T-norms are negative reinforcement operators ($T(x_1, \ldots, x_n) \leq \min_i[T(x_i)]$), but they are not positive reinforced. In addition, the T-conorms are positive reinforced ($C(x_1, \ldots, x_n) \geq \max_i[C(x_i)]$), but they are not negative reinforced. We could hope that combinations of T-norms and T-conorms (as connective mixed ones) are completely reinforced, but Yager and Rybalov found counterexamples proving that this is not true (Yager and Rybalov 1998).

The mean operators are not positively reinforced or reinforced negatively by definition. Indeed, for an average, $\min_i(x_i) \leq M(x_1, \ldots, x_n) \leq \max_i(x_i)$.

The only operator which is (to our knowledge) completely reinforced is the triple Π defined by Yager and Rybalov (1998):

$$PI(x_1, \ldots, x_n) = \frac{\Pi_{j=1}^{n} x_j}{\Pi_{j=1}^{n} x_j + \Pi_{j=1}^{n}(1 - x_j)}. \tag{12.8}$$

Recall that this operator is also a symmetrical sum (Silvert 1979). It is also to be noted that the symmetric sum of two fuzzy sets has the property that the sum of complements is the complement of the sum. We must also note that there are several studies and works on the general properties of symmetrical sums (Silvert 1979; Dubois et al. 1993), but there are few works done on the differences between symmetrical sums.

The property of the total reinforcement is thus particularly interesting because it makes it possible to obtain a good modeling of the human behavior, which is often the goal of many knowledge-based systems. It should be noted that the triple Π incorporates information of the type of the *observations in order to refine the information related to the real world* (Bloch and Hunter 2001) and can be used in this type of information fusion.

12.5 The *Mean Triple Π*

12.5.1 The Mean Triple Π

Although the triple Π is an interesting operator because of the fact that it is completely reinforced, it is sometimes more judicious to use operators of the type *mean*. As first underlined by Yager (1996) and then by Bloch and Hunter (2001) and

Dubois and Prade (2004), when the signals are used to represent the same phenomenon (these signals may be independent of others or not), it is more relevant, from a conceptual point of view, to use a mean operator in order to synthesize the information.

The basic idea which leads to the definition of this new mean operator is to seek a mean operator which has properties close to those of the triple Π. This means that

$$PI(x_1, \ldots, x_n) = \frac{\Pi_{j=1}^{n} G(x_j)}{\Pi_{j=1}^{n} G(x_j) + \Pi_{j=1}^{n} G(1 - x_j))}, \tag{12.9}$$

where $G(x)$ is a function named *generatrix function* which is positive and increasing (Waissman-Vilanova 2000; Silvert 1979). To obtain the idempotent property, we have considered the function $G(x) = x^{1/n}$, where n is the dimension of the vector x. We can define a new aggregation operator,

$$MPI(x_1, \ldots, x_n) = \frac{\Pi_{j=1}^{n} (x_j)^{(1/n)}}{\Pi_{j=1}^{n} (x_j)^{(1/n)} + \Pi_{j=1}^{n} (1 - x_j)^{(1/n)}} \tag{12.10}$$

$$= \frac{1}{1 + \Pi_{j=1}^{n} [\frac{1-x_j}{x_j}]^{1/n}}. \tag{12.11}$$

We call this new operator *mean triple* Π, by reference to the triple Π from which it is obtained.

Proposition 12.1 *The mean triple* Π *defined above is a mean operator.*

Proof We show that this operator is a mean operator by checking the properties of the mean operators (see Yager 1996):

1. the commutativity: $MPI(x, y) = MPI(y, x)$;
2. the monotonicity: $MPI(x, y) \geq MPI(z, t)$ if $x \geq z$ and $y \geq t$;
3. the idempotency: $MPI(x, \ldots, x) = x$;
4. the self-identity : $MPI[B, \langle MPI(B)\rangle] = MPI(B)$.

The first three conditions can be deduced easily from the properties of the product function and n-square function. The most difficult property is the self-identity shown below. We want to demonstrate that:

$$MPI\big(x_1, \ldots, x_n, MPI(x_1, \ldots, x_n)\big) = MPI(x_1, \ldots, x_n, MPI) = MPI(x_1, \ldots, x_n).$$

Therefore,

$$MPI(x_1, \ldots, x_n, MPI) = \frac{\Pi_{j=1}^{n} (x_j)^{1/(n+1)} \times (MPI)^{1/(n+1)}}{D}$$

with

$$D = \left[\Pi_{j=1}^{n} (x_j)^{1/(n+1)} \times (MPI)^{1/(n+1)} + \Pi_{j=1}^{n} (1 - x_j)^{1/(n+1)} \times (1 - MPI)^{1/(n+1)}\right]$$

and

$$(MPI)^{1/(n+1)} = \left(\frac{\Pi_{j=1}^{n}(x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)}} \right)^{1/(n+1)}$$
$$= \frac{\Pi_{j=1}^{n}(x_j)^{1/n(n+1)}}{(\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)})^{1/(n+1)}}.$$

By simplification with

$$\frac{1}{(\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)})^{1/(n+1)}}$$

we have

$$MPI(x_1, \ldots, x_n, MPI)$$
$$= \frac{\Pi_{j=1}^{n}(x_j)^{[1/(n+1)+1/n(n+1)}]}{\Pi_{j=1}^{n}(x_j)^{[1/(n+1)+1/n(n+1)]} + \Pi_{j=1}^{n}(1-x_j)^{[1/(n+1)+1/n(n+1)]}}$$
$$= \frac{\Pi_{j=1}^{n}(x_j)^{[(n+1)/n(n+1)]}}{\Pi_{j=1}^{n}(x_j)^{[(n+1)/n(n+1)]} + \Pi_{j=1}^{n}(1-x_j)^{[(n+1)/n(n+1)]}}$$
$$= \frac{\Pi_{j=1}^{n}(x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)}}$$
$$= MPI = MPI(x_1, \ldots, x_n). \qquad \square$$

The mean triple Π is a new mean operator of aggregation obtained from the triple Π. It is obvious that the mean triple Π cannot be completely reinforced since by definition it is a mean: the numerical value is between the maximum and minimum. However, it has a property similar to the total reinforcement of triple Π, which is based on a comparison with the classic arithmetic mean. The definition of the property is given below.

12.5.2 The Mean Reinforcement

Property 12.1 *Let MPI be the mean triple Π. We consider the classic arithmetic mean $\frac{1}{n}\sum_{j=1}^{n} x_j$. Then: If $x_j \geq 0.5$, $j \in 1, \ldots, n$, then we have*

$$MPI(x_1, \ldots, x_n) \geq \frac{1}{n}\sum_{j=1}^{n} x_j; \tag{12.12}$$

if $x_j \leq 0.5$, $j \in 1, \ldots, n$, *then we have*

$$MPI(x_1, \ldots, x_n) \leq \frac{1}{n} \sum_{j=1}^{n} x_j. \tag{12.13}$$

This property is called mean reinforcement by reference to the total reinforcement of the triple Π.

Proof First, let $x_j \geq 0.5$ for all $j \in 1, \ldots, n$. We have to show that

$$\frac{\Pi_{j=1}^{n}(x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)}} \geq \frac{1}{n} \sum_{j=1}^{n} x_j.$$

This is equivalent to

$$\begin{aligned}
\frac{\Pi_{j=1}^{n}(x_j)^{(1/n)} + \Pi_{j=1}^{n}(1-x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)}} &\leq \frac{n}{\sum_{j=1}^{n} x_j}, \\
1 + \frac{\Pi_{j=1}^{n}(1-x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)}} &\leq \frac{n}{\sum_{j=1}^{n} x_j}, \\
\frac{\Pi_{j=1}^{n}(1-x_j)^{(1/n)}}{\Pi_{j=1}^{n}(x_j)^{(1/n)}} &\leq \frac{n}{\sum_{j=1}^{n} x_j} - 1, \\
\left(\prod_{i=1}^{n}\left(\frac{1}{x_i} - 1\right)\right)^{\frac{1}{n}} &\leq \frac{1}{\frac{\sum_{i=1}^{n} x_i}{n}} - 1,
\end{aligned}$$

and, finally, by taking the logarithms, to

$$\frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{1}{x_i} - 1\right) \leq \ln\left(\frac{1}{\frac{\sum_{i=1}^{n} x_i}{n}} - 1\right).$$

The second derivative of the function $f(x) = \ln(\frac{1}{x} - 1)$ is

$$\frac{d^2 f(x)}{dx^2} = \frac{1 - 2x}{(x^2 - x)^2}.$$

The inequality above follows, by the Jensen inequality on the interval [0.5, 1], from the concavity of f within interval [0.5, 1], where the second derivative of f is negative: $\frac{d^2 f(x)}{dx^2} \leq 0$.

The second inequality follows similarly from the fact that f is convex on $[0, 0.5]$. □

This property translates the fact that when the signals are in accord, the average triple Π discriminates the classes better that the arithmetic mean. This property

can be interesting when we want to evaluate the correlations between different the sources.

12.6 LAMDA Clustering System

The clustering methods deal with unsupervised classification. The term cluster analysis was coined by Tryon in 1939 that encompasses a number of methods and algorithms for categorizing objects of similar kinds. The main objective of clustering is organize a collection of data items into some meaningful clusters, so that items within a cluster are more similar to each other than they are to items in the other clusters. A class or cluster is a set of similar objects (having similar characteristics). Clustering (or segmentation) of objects starts with an arbitrary choice of a similarity measure that describes proximity between different objects. The choice of the similarity (or dissimilarity/distance) measure ultimately defines the outcome of the clustering and is far more important than the choice of the actual clustering algorithm. In general, any similarity measure can be converted into a dissimilarity measure by applying a suitable decreasing function. Some clustering algorithms assume that the dissimilarity or distance measure is a metric.

LAMDA is a fuzzy methodology of conceptual clustering and classification based on the concept of adequacy to each class that replaces the usual "distance to a center" approach. Moreover, the class adequacy concept is expressed as the "fuzzy" truth value of a compound sentence using logical connectives between elementary assertions. We would like to point out that LAMDA method treats objects in a sequential manner. By its value, each descriptor contributes to the global adequacy of one object to one class through marginal adequacy degree (MAD). We use a fuzzy logic operator which interpolates between union and intersection with an adjustable parameter called "exigency." To make a direct confrontation between classes and objects possible, it is necessary that the concept be described with the same descriptor as the one used for observations. Each object is described by a set of attributes or descriptors and represented by vectors of n components. In LAMDA method, descriptors can be considered qualitative or quantitative.

The main properties of LAMDA are the following:

1. both supervised and unsupervised learning may be carried out,
2. simultaneous processing of numerical and qualitative information,
3. learning is performed in a sequential and incremental manner,
4. classification algorithms are based on linear compensated hybrid connectives which aggregate the marginal adequacy degrees (*MAD*) to obtain the global adequacy degree (*GAD*) of an object to a class,
5. total indistinguishability (chaotic homogeneity) in the description space is modeled by means of a special class called the Non-Informative Class (*NIC*); this class accepts any objects with the same adequacy degree and thus naturally introduces a classification threshold,
6. possibility to obtain different classifications from the same group of objects by means of the "exigency" concept.

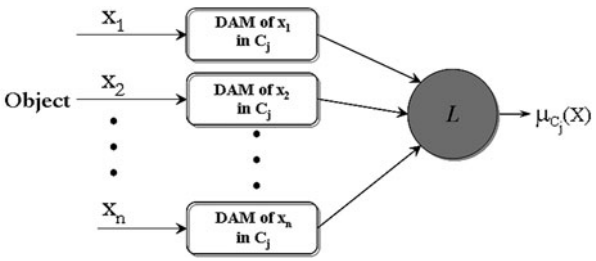

Fig. 12.1 The LAMDA structure

The input data for this algorithm is a collection of objects (individuals or observations), described by a set of n qualitative or quantitative descriptors (attributes), represented as vectors $\mathbf{x}$, where the jth component is the value taken by the jth descriptor if it is numeric or symbolic. In the case of a qualitative descriptor, this value is called modality. The information carried by each descriptor contributes to the membership of the element to the class by means of the Marginal Adequacy Degree (*MAD*) (see Fig. 12.1). LAMDA methodology can be split in 3 stages:

1. Compute the membership function of each source named Marginal Adequacy Degree *MAD*. This membership function is similar to an a posteriori Bayesian probability.
2. The assignment of any object to a class is computed by fusion of all marginal information available *MAD* using a fuzzy fusion operator, the result being named *GAD* (Global Adequacy Degree). We remark that $MAD = 1$ represents the total adequacy of the given attribute to the class and $MAD = 0$ represents its total inadequacy.
3. Whenever the *NIC* class corresponds to the maximum *MAD*, the object is considered as unrecognized, and no known concept is related to it. Two alternatives exist:
   a. the object is merely ignored, or unclassified;
   b. the object is considered to belong to an unknown concept, and it will be taken as the first element of a new class. This case is called self-learning. The new concept (class) will be initialized by this object and by the parameters of the *NIC*.

The marginal adequacy degree function, chosen in this application, is a fuzzy interpolation of the binomial probability:

$$MAD_{j,k} = \rho_{j,k}^{\delta(x_j, c_{j,k})} (1 - \rho_{j,k})^{1-\delta(x_j, c_{j,k})}, \tag{12.14}$$

where $\delta(x_j, c_{j,k})$ is a distance toward a central parameter of class C_k. To estimate the parameters $\rho_{j,k}$ and $c_{j,k}$ for a given learning data set, the minimization of a likelihood criterion is used:

$$J(\rho, c) = \max_{\rho,c} \left| MAD(x_i)_{i=1}^{n} \right|. \tag{12.15}$$

This is obtained for each descriptor (the index of descriptor is omitted) by

$$\rho_k = \frac{1}{n} \sum_{i=1}^{n} \sigma(x_i, c_{i,k}) \tag{12.16}$$

and resolving

$$\sum_{i=1}^{n} \frac{\partial}{\partial c_{i,k}} \delta(x_i, c_{i,k}) = 0. \tag{12.17}$$

The distance used here is the scalar Euclidean $\delta(x_i, c_{i,k}) = |x_i, c_{i,k}|$.

It must be noted that in order to calculate the adequacy of an element to a class, both must have the same descriptor set. Then, all the *MAD* are aggregated in order to obtain a Global Adequacy Degree (*GAD*) of the object to the class. This is made by a convex interpolation of fuzzy logic connectives L_α the Mixed Connective of linear compensation, presented before, which in the new notation become

$$\begin{aligned} &GAD(MAD_1, MAD_n) \\ &\quad = \alpha * T(MAD_1, MAD_n) + (1 - \alpha) * C(MAD_1, MAD_n), \end{aligned} \tag{12.18}$$

where MAD_i is the marginal adequacy of the object and $\alpha \in [0, 1]$, to be coherent with Fuzzy Logic aspects that include compatibility with Boolean Logic; T is an iterated T-norm, and C its dual T-conorm with respect to the negation (complement to 1). The parameter α is called the *Exigency Index*, and it is possible to associate different classifications to the same data set, depending on the value chosen for α. As shown in Piera-Carreté et al. (1988), recognition is more exigent as α increases; therefore, there will be more nonrecognized objects. Similarly, if α increases, learning becomes more selective (or exigent) as the number of objects assigned to the *NIC* increases, and so does the number of created classes. Thus, by changing the value of α, different partitions from the same data set, based on the same logical criterion, can be obtained.

The clustering algorithm LAMDA is presented below:

1. Get the extremal values $x_{i,\min}$ and $x_{i,\max}$ and those of the quantitative components. Replace the value x_i by its normalized value for each descriptor i:

$$x_i = \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}}. \tag{12.19}$$

2. Compute the marginal adequacy degrees for each descriptor which are respectively MAD_i, $i = 0, \ldots, d$.
3. Compute the global adequacy. Search for the maximum *GAD* degree to assign the object to a class.
4. Two cases may occur:
   a. Recognition mode. Object x is placed in C_i. If C_i is *NIC*, then object x is said to be "unrecognized."

b. Self-learning, or Concept Formation mode. There are two possibilities:
   i. GAD_i does not correspond to *NIC*; in this case, x is placed in C_i, and the parameters of C_i will be modified to include x.
   ii. GAD_i corresponds to *NIC*. This means that x is the first element of a new class C_{k+1}, and the representation of this new class will depend on x.

For the modification of class parameters, we use the following algorithm:

$$c_{i,k} = c_{i,k} + \frac{x_i - c_{i,k}}{N_{NIC} + 1} \tag{12.20}$$

and

$$\rho_{i,k} = \rho_{i,k} + \frac{\delta(x_i, c_{i,k}) - \rho_{i,k}}{N_{NIC} + N_k + 1}, \tag{12.21}$$

where N_k is the number of elements assigned to class C_k, and N_{NIC} is a virtual number of elements of the *NIC* class; it is a parameter introduced in order to initialize the new classes whenever self-learning is applied. It can be noticed that $\rho_{i,NIC} = \frac{1}{2}$, so that the initial parameters of the new class are

$$\rho_{i,k} = \frac{1}{2}\big(N_{NIC} + \delta(x_i, c_{i,k})\big)\frac{1}{N_{NIC} + 1} \tag{12.22}$$

and $c_{i,k} = x_i$.

12.7 Experimental Results

Biological knowledge is inherently incomplete, owing to the complexity of living systems and the limitation of scientific methods available for the study of those systems. The incompleteness of knowledge constantly manifests itself unexplained observations. To account for these novel observations, biologists need to revise or extent the existing knowledge. The application that we treat relates to fusion of information during a processes of classification in a biotechnology. We need to fusion information resulting from various measured biochemical parameters (pH, dioxygene, carbon dioxide, etc.) allowing one to carry out a nonsupervised classification. This classification is based on the hypothesis that the measurements expressed the same biological phenomena. Therefore, the obtained classes must correspond to the physiological states of these microorganisms (note that we have *information from real world*). The physiological state is the biological reaction inside the microorganisms which has, for consequence, the production of a specific metabolite or the reproduction of the cells. In this application the physiological states are known based on the analysis of respiratory quotient. The goal of the nonsupervised classification is to determine the class corresponding to the physiological states without any knowledge. We have noticed that the Mixed Connective leads to dissatisfaction in the classification processes. Therefore, we have replaced the Mixed Connective by triple Π and the Mean triple Π.

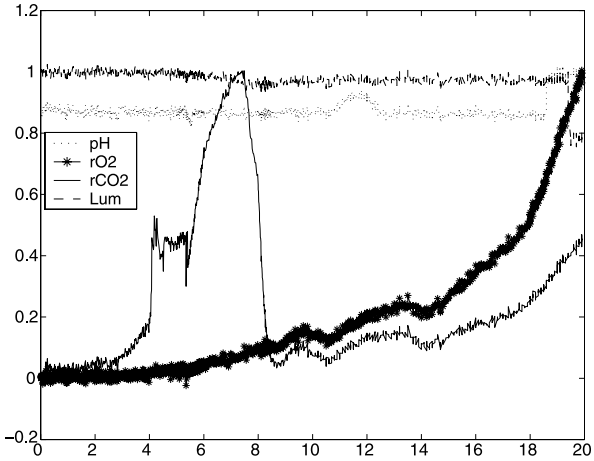

Fig. 12.2 The four biochemical parameters with noise (SNR = 40 dB)

Table 12.1 Comparison of the classification using the mean triple Π (*MPI*) and the triple Π (*PI*). CTN signifies that the Classification is Too Noisy to obtain a significant result

	% classif. M3PI	% classif. 3PI
SNR = 46.02 dB	13.47%	CTN
SNR = 40 dB	11.56%	CTN
SNR = 30.40 dB	CTN	CTN

In comparison to triple Π, the Mean triple Π provides a ranging value between the minimum and the maximum of the marginal membership degrees, making the synthesis between the various manifestations of the same event. In practice, we noted that the results of classification of the two operators were generally similar.

A notable exception between triple Π and the Mean triple Π is the classification of the noisy signals. The mean triple Π due to its property of smoothing is more robust with the noisy signal than triple Π. On Fig. 12.2, four parameters (pH, rO_2, rO_2, and Luminance) have been used. The two operators were tested on these noisy signals, and two classifications are disturbed by the presence of noise. Nevertheless, even in the presence of noise, the classification using the average triple Π provides at least a class which characterizes the fermentation state (state 1); see Table 12.1.

12.8 Uncertainties and Maximum of Modulus of Wavelet Transform

By uncertainties we mean that a transition between two classes is not well defined (a class overlaps the other) or that an isolated point belonging to a class is located among a set of other points belonging to another class. These uncertainties come from the perturbations due to measurement noise or from the analyzed image itself.

The classification is perturbed, and the borders between classes are not always well defined. Moreover, the lack of meaningful data reinforces such uncertainties in classification. To deal with the noise problem, one can use filters, but the result is linked to the nature of the filter, and we do not really know if the filtering causes the loss of meaningful and pertinent information. That is why we propose to use the maximum of modulus of wavelet transform.

12.9 Classification and Maximum of Modulus of Wavelet Transform

12.9.1 The Continuous Wavelet Transform

Wavelet theory is experiencing an increasing success, and wavelets are now being used in many fields. The wavelets have time-scale properties that are very interesting for the analysis of nonstationary signals. As we said above, we suppose that the singular points or the inflexion points may correspond to the transition between two classes and, consequently, between two different regions. To detect those points, we use the Maximum of Modulus of Wavelet Transform (Mallat 1991; Mallat and Zhong 1992). The main particularity of this Maximum of Wavelet Transform is using wavelet which is the first or second derivative of smoothing function (Gaussian for example):

$$\psi(t) = \frac{d\theta(t)}{dt}. \tag{12.23}$$

The wavelets are a powerful mathematical tool of nonstationary signal analysis (of signals whose frequencies change with time). Unlike the Fourier Transform, Wavelet Transform can provide the time-scale localization. The performance of the Wavelet Transform is better than that of the windowed Fourier Transform. Because of these characteristics, Wavelet Transform can be used for analyzing nonstationary signals such as transient signals. Wavelets Transformation (WT) is a rather simple mechanism used to decompose a function into a set of coefficients depending on scale and location. The definition of the Wavelets Transform is

$$W_{s,u} f(x) = (f \star \psi_{s,u})(x) = \int f(x)\psi\left(\frac{x-u}{s}\right)dx, \tag{12.24}$$

where ψ is the wavelet, f is the signal, $s \in R^{+*}$ is the scale (or resolution) parameter, and $u \in R$ is the translation parameter. The scale plays the role of frequency. The choice of the wavelet ψ is often a complicated task. We assume that we are working with an admissible real-valued wavelet ψ with r vanishing moments ($r \in N^*$).

The wavelet is translated and dilated as in the following relation:

$$\psi_{u,s} = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{12.25}$$

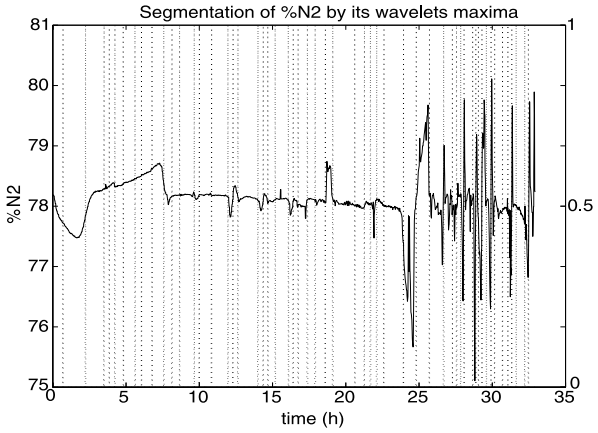

Fig. 12.3 Segmentation of N2 (nitrogen). Each vertical dotted line corresponds to a singularity of the signal detected by wavelets. The wavelet is a DOG (the first derivative of Gaussian), and the scales go from 2^0 to 2^3

The dilation allows the convolution of the analyzed signal with different sizes of "window" wavelet function. For the detection of the singularities and of the inflexion points of the biochemical signal, we use the Maxima of Modulus of Wavelets Transform (Mallat and Hwang 1992). The idea is to follow the local maxima at different scales and to propagate from low to high frequencies. These maxima correspond to singularities, particularly when the wavelet is the derivative of a smooth function,

$$\psi(x) = \frac{d\theta(x)}{dx},$$

$$W_{s,u} f(x) = f * \psi_{s,u} = f(x) * \frac{d\theta(x/s)}{dx}.$$

Yuille and Poggio (1986) have shown that if the wavelet is the derivative of the Gaussian, then the maxima belong to connected curves that are continuous from one scale to another. The detection of the singularities of the signal is thus possible by using the wavelets (see, for example, Fig. 12.3).

The discretization form of Continuous Wavelet Transform is based on the following form of the Mother Wavelet:

$$\psi^{m,n}(t) = a_0^{-m/2} \psi\left(\frac{t - nb_0 a_0^m}{a_0^m}\right). \tag{12.26}$$

By selecting a_0 and b_0 properly, the dilated mother wavelet constitutes an orthonormal basis of $L^2(R)$. For example, the selection of $a_0 = 2$ and $b_0 = 1$ provides a dyadic-orthonormal Wavelet Transform (DWT). The decomposed signals by *DWT* will have no redundant information thanks to the orthonormal basis.

Jiang et al. (2003) have proposed to select the maxima by using thresholding. Besides, all the singularities are not relevant; only some of them are meaningful.

However, as stated above, the thresholds proposed by Jiang et al. are chosen empirically. To select the meaningful singularities, we propose using the Hölder exponent. The Hölder exponent is a mathematical value allowing characterization singularities. The fractal dimension could also be used, but only the Hölder exponent can characterize locally each singularity. A singularity at a point x_0 is characterized by the Hölder exponent (also called the Hölder coefficient or Lipschitz exponent). This exponent is defined as the most important exponent α allowing us to verify the inequality

$$\left|f(x) - P_n(x - x_0)\right| \leq C|x - x_0|^{\alpha(x_0)}. \tag{12.27}$$

We must remark that $P_n(x - x_0)$ is the Taylor Development and basically $n \leq \alpha(x_0) < n + 1$. The Hölder exponent measures the remainder of a Taylor expansion and, moreover, measures the local differentiability:

1. $\alpha \geq 1$: $f(t)$ is continuous and differentiable.
2. $0 < \alpha < 1$: $f(t)$ is continuous but nondifferentiable.
3. $-1 < \alpha \leq 0$: $f(t)$ is discontinuous and nondifferentiable.
4. $\alpha \leq -1$: $f(t)$ is no longer locally integrable.

Therefore, the Hölder exponent can be extended to a distribution. For example, the Hölder exponent of the Dirac function is equal to -1. A simple computation leads to a very interesting result of the Wavelets Transform (Jaffard 1997):

$$\left|W_{s,u} f(x)\right| \simeq s^{\alpha(x_0)}. \tag{12.28}$$

This relation is remarkable because it allows one to measure the Hölder exponent using the behavior of the Wavelets Transform. Therefore, at a given scale $a = 2^N$, the $W_{a,b} f(x)$ will be the maximum in a neighborhood of the signal singularities. The detection of the Hölder coefficient is linked to the vanishing moment of the wavelet: if n is the vanishing moment of the wavelet, then it can detect the Hölder coefficients less than n (Mallat and Hwang 1992). We use a *DOG* wavelet (DOG: the first derivative of Gaussian) with a vanishing moment equal to 1; consequently, we can only detect the Hölder coefficients smaller than 1. This is not a real problem because we are interested (*in this application*[1]) by the singularities such as step or Dirac, and the Hölder coefficients of these singularities are smaller than 1. The values of the used integers are not meaningful: they are consecutive, and the only obligation is that an integer corresponds to one and only one class. The functions by stage are introduced in the classification as a new descriptor of the object. Besides, the maximum of modulus wavelet transform and, obviously, the functions by stage represent the borders of the different regions in the images. The functions by stage make it possible to have more precise borders in the LAMDA classification, as we will see in the next section.

[1]However, for others applications in bioprocesses, it is always possible to use other wavelets with greater vanishing moments.

12.9.2 Maximum into the Classification

The function by stage enables us to influence the classification of LAMDA. Because of the fact that the *GAD* triple Π is a total reinforced operator (Yager and Rybalov 1998) (unlikely the T-norm which is only negatively reinforced), LAMDA tends to facilitate the functions by intervals in comparison with the other data (here, the biochemical parameters) providing that these functions by stages are meaningful. Let us demonstrate this assertion.

Proof The triple Π is a reinforced operator, that is, it is positively reinforced and negatively reinforced. Particularly, the triple Π is positively reinforced, i.e., if all the descriptors are affirmative (that is, all the *MAD* are higher than 0.5), then we have

$$GAD_{j,i}(x_1, \ldots, x_n) \geq \max_{i=1,\ldots,n} \left[MAD_{j,i}(x_i)\right] \tag{12.29}$$

□

Let us take one meaningful stage of the function by stage, this function by stage being the lth descriptor of the objects to classify, and the integer distinguishing this stage being noted A. The class featuring by this particular stage is called C_s. This stage is meaningful, and it agrees with the *MAD* of the line of level gray in the temporal interval where the stage is defined. Let us suppose that all those *MAD* are *affirmative*. For all the objects that are located in the temporal interval where the stage is defined, the *MAD* for this stage is maximum. Besides, for an object located in this interval, the distance between the lth descriptor of the object which corresponds to this function by stages and the lth component of the center (called $c_{s,l}$) is equal to zero as we have

$$\alpha(x_l, c_{s,l}) = \alpha(A, A) = 0. \tag{12.30}$$

By consequence the *MAD* for the descriptor l is maximum:

$$MAD_s(x_l) = MAD_{s,l} = \rho_{sl}^{1-\alpha(x_l, c_{s,l})}(1 - \rho_{sl})^{\alpha(x_l, c_{s,l})} = \rho_{sl}. \tag{12.31}$$

Thus, according to (12.29), the *GAD* will tend to be very important for this class. In fact, the *GAD* will be equal to one because of the value of ρ_{sl} (one can easily show that here $\rho_{sl} = 1$ according to (12.16)) as we can see in the following equation:

$$MAD_{s,l} = \rho_{sl} = 1;$$

then

$$GAD_s(x_1, \ldots, x_l, \ldots, x_n) \geq MAD_{s,l} = 1;$$

and then

$$GAD_s(x_1, \ldots, x_l, \ldots, x_n) = 1.$$

If we compute the *MAD* for the other stages (more precisely, for the classes featuring by those classes $C_{r,r\neq s}$), they will not always be as optimal as the *MAD* of the class C_s since the values characterizing those stages are all different of A, and thus the distance between the lth component of the object and the lth component of the center of each class will always be greater than zero. For those classes, the *MAD* will not be as important as for the class C_s, i.e.,

$$GAD_{r,r\neq s}(x_1,\ldots,x_l,\ldots,x_n) \leq 1 = GAD_s(x_1,\ldots,x_l,\ldots,x_n).$$

So the maximum provides implicitly the borders of the classes and thus helps the region detection.

12.10 Conclusion

According to the analysis made up in the first section of the chapter, for the sources of information describing the same phenomena, it is better to use a mean operator. The results of nonsupervised time series classification obtained using 3Π and mean triple Π have shown that there are no revelent differences between these operators. The explication is based on the fact that the sources are in perfect concordance; therefore, the reinforced operator and the mean operators give very close resulting scores. Of course, the new operator of aggregation mean triples Π seems well adapted to analyze the time series describing the strong nonlinearity of biological systems. This operator combines the properties of the completely reinforced operators and the mean operators. The results obtained on the time series show that this operator is less sensitive to the noise than the operator of Yager and Rybalov (1998). Therefore, in the case of biological signals which have an important response time and are noisy, we can use it as an aggregation operator. In the future, we want to compare the classification obtained by triple Π and mean triple Π in the goal of introducing the notion of pertinence of data in biological system analysis.

References

I. Bloch, Information combination operators for data fusion: a comparative fusion with classification. Technical report, ENST, Paris, 1994. Rapport Technique.

I. Bloch and A. Hunter, Fusion: general concepts and characteristics. *International Journal of Intelligent Systems*, 16:1107–1134, 2001.

T. Calvo, B. De Baets, and J. Fodor, The functional equations of Frank and Alsina for uninorms and nullnorms. *Fuzzy Sets and Systems*, 120:385–394, 2001.

D. Dubois and H. Prade, La fusion d'informations imprécises. *Traitement du Signal*, 11(6):447–458, 1994.

D. Dubois and H. Prade, On the use of aggregation operations in information fusion process. *Fuzzy Sets and Systems*, 142:143–161, 2004.

D. Dubois, H. Prade, and R. R. Yager, editors, *Readings in Fuzzy Sets for Intelligent Systems*. Morgan Kaufman, San Mateo, 1993.

S. Jaffard (1997), Multifractal formalism for functions part 1 and 2. *SIAM Journal of Mathematical Analysis*, 28(4):944–998.

T. Jiang, B. Chen, X. He, and P. Stuart (2003), Application of steady-state detection method based on wavelet transform. *Computer and Chemical Engineering*, 27(4):569–578.

S. Mallat, Zero crossing of a wavelet transform. *IEEE Transactions on Information Theory*, 37:1019–1033, 1991.

S. Mallat and W.-L. Hwang (1992), Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2):617–643.

S. Mallat and S. Zhong (1992), Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710–732.

N. Piera-Carreté, J. Aguilar-Martin, and M. Sanchez, Mixed connectives between min and max. In *8th Inter. Symp. on Multiple Valued Logic*, 1988.

W. Silvert, Symmetric summation: A class of operations on fuzzy sets. *IEEE Transactions on Systems, Man, Cybernetics*, 9(10):657–659, 1979.

J. Waissman-Vilanova, *Construction d'un modèle comportemental pour la supervision de procédés: application à une station de traitement des eaux*. PhD thesis, LASS–CNRS, Novembre 2000.

R. Yager, On ordered weighted averaging operators in multi-criteria decision making. *IEEE Transactions on Systems, Man, Cybernetics*, 183–190, 1988.

R. Yager, On mean type aggregation. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 26(2):209–221, 1996.

R. Yager and A. Rybalov, Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80(1):111–120, 1996.

R. Yager and A. Rybalov, Full reinforcement operators in aggregation techniques. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 28(6), 1998.

A. Yuille and T. Poggio (1986), Scaling theorems for zero-crossing. *IEEE Transaction for Zero-Crossing*, 8(1):15–25.

H. Zimmerman and P. Zynso, Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.