

**The logical combinatorial approach to Pattern Recognition, an overview through  
selected works**

**J. Francisco Martínez-Trinidad\* and Adolfo Guzmán-Arenas**

**Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City  
{aguzman, fmartine}@cic.ipn.mx**

*ABSTRACT.* The so-called *logical combinatorial approach* to Pattern Recognition is presented, and works (mainly in Spanish and Russian) that are not ordinarily available, are exposed to the Western reader. The use of this approach for supervised and unsupervised pattern recognition, and for feature selection is reviewed. Also, an unified notation describing the original contributions is presented, thus rendering this important area more readable.

Our review is not exhaustive; nevertheless, most significant works are enclosed. Our hope is to motivate the reader to inquire further in these works.

This paper serves as an introduction to three articles on the *logical combinatorial approach* that appear in this issue of *Pattern Recognition*.

*KEYWORDS:* *Pattern recognition, classification, feature selection, testor theory, logical combinatorial Pattern Recognition.*

## **INTRODUCTION**

---

\* Centro de Investigación en Computación del IPN, P. O. Box 75-476, 07738 Mexico City, MEXICO.  
fmartine@cic.ipn.mx

Several tools have been proposed since the beginning of Pattern Recognition, to solve three important problems: feature selection, supervised classification, and unsupervised classification or clustering. Fukunaga [1] defines *feature selection for representation* as a mapping from the original features (or variables) into those most efficient; tools are used for this purpose such as the discrete Karhunen-Loève expansion, the same expansion for random processes, the estimation of eigenvalues and eigenvectors, principal component analysis [2], and factor analysis [3]. Also *Feature selection for classification* is defined in [1] as: given two or more classes, select those features that are most efficient to preserve class separability. To solve this problem, one resorts to discriminant analysis, non parametric discriminant analysis, sequential selection of quadratic variables, as well as several selection criteria of subsets of features that optimize some separability criteria among classes.

In *supervised classification*, the problem is to recognize, given a set of objects grouped into classes, in which one (or more than one) of these classes new objects or measurements belong. Several classifiers have been proposed: maximum likelihood; Bayesian, (1-nn), k nearest neighbors (k-nn), linear, quadratic, and other [1, 4-6].

In *unsupervised classification*, a sample of objects is at hand, but their clustering into groups or classes is unknown. The problem consists in defining such classes. To solve these problems, tools such as ISODATA [7], C-means [8], unsupervised Bayesian classifiers [6], grouping and dividing hierarchical classifiers [9], graph-theoretic classifiers [10, 11] are used, among others.

Work mentioned so far falls under Statistical Pattern Recognition, characterized because it works with the descriptions of the objects under study, contrasting with syntactic structural recognition [4], which works with the *parts* of the objects under study. This paper reviews

several articles developed under the Logical Combinatorial Pattern Recognition approach, in the three areas above mentioned. This approach, as the statistical approach, works with the descriptions of the objects. In this issue, a paper [12] introduces the concepts of testor, typical testor, logical-combinatorial approach, fuzzy testor, and others.

Let  $U$  be a (perhaps infinite) universe of objects, and, following the statistical approach, let us consider a given finite sample  $M=\{O_1,\dots,O_m\}$  of such (descriptions of the) objects. We shall denote by  $R=\{x_1,\dots,x_n\}$  the set of features or variables used to study these objects. Each of these features has associated a set of admissible values (its domain of definition)  $M_i, i=1,\dots,n$ . These sets of values, in contrast to the other approaches, can be of any nature: variables can be quantitative and qualitative simultaneously. Each of these sets contains a special symbol denoting absence of information (missing data). Thus, some variables are numeric; other, symbolic; incomplete information about some objects is allowed. This will turn out to be a fundamental feature of this Pattern Recognition paradigm. By a description of an object  $O$  we understand an  $n$ -tuple  $I(O)=(x_1(O),\dots,x_n(O))$  where  $x_i:M\rightarrow M_i$ , for  $i=1,\dots,n$  are the variables or features used to describe it. Over  $M_i$  no algebraic, topologic or logic structure is assumed. Any of the Pattern Recognition problems mentioned above is formulated from a set of descriptions of  $m$  such objects.

## **FEATURE SELECTION**

Let  $M$  be a given set of descriptions of  $m$  objects from  $U$  in terms of the features of  $R$ . We assume that the descriptions of  $M$  are structured in  $r$  subsets (classes)  $K_1,\dots,K_r$  (not necessarily disjoint, not necessarily crisp). Under these conditions two kinds of problems can be formulated: feature selection for classification; same for description. The first case

tries to find the subset(s) of features that determine representation subspaces of the objects which are most appropriate for a later problem of supervised classification. In the second case the objective is not classification of new objects, but to determine the informational relevance of each of the features, or of subsets of them. Feature selection in this field was initiated by Dmitriev, Zhuravliov and Krendeleiev [13]. A recent book [25] goes further into this area.

Ortiz-Posadas [14] applies the logical-combinatorial model of Pattern Recognition Theory [15] to the computer-assisted medical diagnostic and prognostic. She uses three models: a medical model, of Heathfield *et al.* [16], for histopathologic diagnosis of breast diseases; a mathematical model, provided by the logical-combinatorial model of Pattern Recognition; and a computational model. Her dissertation explains how to select a suitable set of variables for the case in question. Some medical results appear in [17-18].

A CINESTAV report [19] presents an extension to *typical testor* using analogy between patterns, and an algorithm to detect *fuzzy typical testors* of a training matrix.

A short letter [20] extends the expression to determine the feature relevance (that is, how relevant or discriminant a given feature of an object is) in crisp and fuzzy environments, and introduces an algorithm (different to that of [19]) to compute all the fuzzy typical testors of a training matrix.

Carrasco-Ochoa [21] analyzes how changes in the training matrix affect its typical testors. He does this for three kind of typical testors: (1) those of Zhuravlev [22, 23]; (2) the  $\epsilon$ -typical testors, which use a similarity function between two objects that makes two objects similar if they differ in  $\epsilon$  (an integer  $\geq 0$ ) or less feature values; (3) those of Goldman [24].

Lazo-Cortés *et al.* [12], in this issue, give several contrasting definitions of *testor*, and their application in supervised classification. This introduces the *logical combinatorial approach* to Pattern Recognition. An excellent introduction is also in a book [25].

Ozols and Borisov [26], in this issue, introduce *shadows composition* in order to solve fuzzy supervised classification.

Martínez-Trinidad and Ruiz-Shulcloper [27], in this issue, use the fuzzy clustering criteria of the logical-combinatorial approach, to do unsupervised pattern classification of non-numerical terms belonging to a vocabulary.

## **SUPERVISED CLASSIFICATION**

In this kind of problems, we assume that the universe  $U$  is structured in a finite number  $K_1, \dots, K_r$  of proper subsets, called classes, and from each of them we have a sample of descriptions of objects, the so-called training matrix  $MA = \{ K_1 \cup, \dots, \cup K_r \}$ . The problem is to find the membership relations from a new object from  $U$  (outside the given samples) with the  $r$  classes. This relationship does not have to be all or nothing. The logical combinatorial approach deals with spaces without algebraic (or of any other kind of) structure. The representation space is simply a Cartesian product, which also has the peculiarity of being heterogeneous, that is, each of the sets forming it can be of different nature: a set of real numbers, a set of labels, a set of truth values from a given logic, etc. An example of this appears in medical diagnosis problems, where descriptions take the form  $I(O) = (\text{black, female, 45, } 38.6^0, 1500, ?, \text{slight, } \dots)$ , where ? means absence of information. In other disciplines, too, such as Geo Sciences, Sociology, Pedagogy, Marketing, etc. (the so called

soft sciences), objects are described in terms of qualitative and quantitative variables. Thus, the tools herein presented.

Most significant models of supervised classification are those works based on *partial precedences*. There are three models: voting algorithms, Kora model, and algorithms based on representative sets.

### **Voting algorithms**

These were conceived in a simplified manner by Zhuravlev and his group [28]. The main idea is partial precedence, that is, partial analogies. An object can resemble another object not in its totality; those parts which *do* resemble each other can give information about possible regularities; of course not all of the same magnitude. For practical problems in the disciplines cited above, this model was very limited, since it only allows object descriptions based in Boolean variables. A later paper [25, 29] added some results, producing an improved parametric model comprising six steps: 1) defining the system of support sets; 2) defining the similarity function; 3) row evaluation, given a fixed support set; 4) class evaluation for a fixed support set; 5) class evaluation for all the system of support sets, and 6) resolution rule. Thus, to define a voting algorithm, is to define a set of parameters for each of the above six steps.

A support set is a non empty subset  $\Omega = \{x_{i_1}, \dots, x_{i_s}\}$  of features which shall be used to analyze the objects. We denote as  $\Omega O$  the subdescription in terms of the features of  $\Omega$ . Thus, a system of support sets denoted by  $\Omega_A$  are several support sets which together will allow analysis of the objects to be classified, comparing them with objects in each one of the classes  $K_i$   $i=1, \dots, r$ . Note that said analysis is done paying attention to different parts or

subdescriptions of the objects, and not analyzing the complete descriptions. Examples of systems of support sets are the set of typical testors, combinations with a fixed cardinality, combinations with variable cardinality, the power set of features, etc.

A similarity function defines the form to compare the descriptions (subdescriptions) of the objects. Ruiz-Shulcloper and his group [25, 29] propose to do this starting by determining the comparison among the values of features, and evaluating these comparisons through a similarity function. A simple example of similarity function is

$$\Gamma(\Omega O_i, \Omega O_j) = \sum_{x_p \in \Omega} C_p(x_p(O_i), x_p(O_j)).$$

When the systems of support sets and the similarity function have been defined, the voting process starts in the stage of row evaluation; that is, the similarity between the different parts of the objects already classified and those to be classified is analyzed. Each row of  $MA$  (each object  $O_i \in MA$ ) is compared with object  $O$  to be classified using the similarity function  $\Gamma$ . This evaluation is a function of the similarity values among the different parts being compared. An example of this evaluation is  $\varphi_\Omega(O_i, O) = \gamma(O_i)P(\Omega)\Gamma(\Omega O_i, \Omega O)$  where we consider a weight  $\gamma(O_i)$  associated to each object  $O_i$  from  $MA$  and a weight  $P(\Omega)$  for the support set  $\Omega$ .

The class evaluation for a fixed support set  $\Omega$  consists in totaling the evaluations obtained for each of the objects  $MA$  with respect to the object  $O$  to be classified. This total evaluation is a function of the row evaluations already obtained. An example of this evaluation is:

$$\varphi_\Omega^j(O) = \frac{1}{|K_j|} \sum_{t=1, \dots, |K_j|} \varphi_\Omega(O_t, O), \text{ the upper index refers to the class } K_j.$$

In the class evaluation for all the system of support sets, evaluations are totaled for all the system of support sets. Following our example, this step could be expressed as follows:

$$\varphi^j(O) = \frac{1}{|\Omega_A|} \sum_{\Omega \in \Omega_A} \varphi_{\Omega}^j(O).$$

Finally, the resolution rule is a function that establishes a criterion taking into account each voting thus obtained, and reaches a decision concerning the relations of the object to be classified with every class of the posed problem. In general, this function has the following form:  $r(\varphi^1(O), \dots, \varphi^r(O)) = (\alpha_1(O), \dots, \alpha_r(O))$ . A similar manner to compute the values of

$$\alpha_i(O) \quad i=1, \dots, r., \text{ can be: } \alpha_i(O) = \begin{cases} 1 & \text{if } \varphi^i(O) > \varphi^j(O) \forall i \neq j \\ 0 & \text{otherwise} \end{cases} .$$

This model for supervised classification has been applied to several practical problems [18, 30-33].

An extension to the above model is proposed by Lazo [34], where he lets the features be numeric or linguistic; that is, values can be some labels or words of natural languages (fuzzy variables). The similarity between values of the same feature is evaluated in the interval [0,1], considering as a special case a two-valued comparison criteria. An important element of this new model is that it covers problems considering fuzzy membership (in Zadeh's sense [35]) of the objects to the classes. In this manner, he can work with fuzzy support sets in which each feature belongs to a certain degree to the support set. For instance, fuzzy Goldman testors can be used as support sets. The evaluation of the similarity between objects, proposed in this work, is summarized in the following general expression:  $\Gamma(\Omega O_i, \Omega O_j) = |\Omega|^{-1} \sum_{x_p \in \Omega} \mu_{\Omega}(x_p) C_p(x_p(O_i), x_p(O_j))$ . This expression considers

the degree of membership of feature  $x_p$  to the support set  $\Omega$ . In addition, if  $\Omega$  is a fuzzy set, the scalar cardinality (see Zimmermann [36]) is considered. In the case of linguistic

variables, one can use as criteria for comparing values of features the Hamming, Euclid or Kolmogorov-Fomin criteria, which are obtained from the homonymous distances between fuzzy sets, see [37]. To evaluate by class using a fixed support set, one works with

$$\varphi_{\Omega}^j(O) = \frac{1}{|K_j|} \sum_{t=1, \dots, |K_j|} \varphi_{\Omega}(O_t, O),$$

if in the formulation of the problem each object belongs to only one class. To consider degrees of membership of each object to each class, the

$$\text{expression } \varphi_{\Omega}^j(O) = \frac{1}{|K_j|} \sum_{t=1, \dots, |K_j|} \mu_{K_j}(O_t) \varphi_{\Omega}(O_t, O) \text{ is used, where } |K_j| \text{ denotes the scalar}$$

cardinality of class  $K_j$ , and  $\mu_{K_j}(O_t)$  denotes the degree of membership of object  $O_t$  to class

$K_j, j=1, \dots, r$ . For evaluation by class for all the system of support sets, one considers the

average of the evaluations, in similar manner as the former model. Finally, in the resolution

rule can be considered each  $\varphi^j(O) j=1, \dots, r$ , for each class as the fuzzy membership degrees

to the classes, or as the votings for belonging to each class.

Finally we can mention the voting algorithm for linguistic variables [34]. In this model the

problem of classification is posed in similar terms to defining the testors to certain degree,

that is, considering that objects belong with certain degree to each class. Similarity

functions are defined between object descriptions in terms of linguistic variables, as well as

expressions for row evaluation of a fixed support set and for all the system of support sets.

### **Kora Model**

These models start with Bongard's group [38], who propose the Kora-3 algorithm for

solution of Geophysics and Geology problems. This was the first algorithm to be used for

solving supervised classification in the Geological environment.

Kora-3 rests on the idea that classes are formed by objects fulfilling certain complex properties, composed by the conjunction of three simple properties. This algorithm is very limited since it only allows Boolean description of objects, and works only with two non-intersecting classes. The basic idea is partial precedence (in similar form to the voting algorithms), but restricted to the use of relations of any three features. A subset of three features  $\Omega = \{x_{i_1}, x_{i_2}, x_{i_3}\}$  and a combination of values  $(a_1, a_2, a_3)$  for  $\Omega$ , form a *complex feature* of class  $K_i$  if and only if the triplet  $(a_1, a_2, a_3)$  appears at least  $\delta_i$  times in the subdescriptions corresponding to the objects of  $K_i$  and does not appear in the subdescriptions of objects of the other class. Those objects in which this combination appears are called *objects characterized* by the complex feature. On the other hand, objects characterized by less than  $\delta_i$  complex features form *the remainder of class  $K_i$* . In this manner, if two complex features  $r_1$  y  $r_2$  characterize exactly the same objects, they are called equivalent, and if  $r_1$  characterizes all objects characterized by  $r_2$  and at least one more, then it is said that  $r_1$  is stronger than  $r_2$ . Based on these concepts the Kora-3 algorithm is defined in three steps: (1) learning step; (2) re-learning step; (3) classification step. In the learning step the complex features are computed for each class using parameters  $\delta_1$  and  $\delta_2$ . In the re-learning step *complementary complex features* are computed using new thresholds  $\delta'_1 < \delta_1$  and  $\delta'_2 < \delta_2$  in the remainder of the classes. Finally, the classification step counts how many complex features of each class characterize or vote in favor of the new object to be classified, and selects the class that provides the largest number of complex features. This algorithm became very limited given the characteristics of practical problems, hence the first extension proposed [39, 40] modified the concept of a complex feature, allowing now a combination of three values to appear  $\delta_i$  times (sufficiently) in the classes and  $\delta'$  times

(sufficiently little) in the other class. Later, De la Vega-Doria *et al.* [41] propose an extension which allows the use of any system of support sets and not only support sets formed by three features, it is allowed to work with more than two classes, overlap among classes is allowed, the type of variables can be of any sort, lack of information is allowed, and in the most general case, classes can be fuzzy, that is, object belong to a class to a certain degree. To reach this extension the concept of complex feature was modified once more, giving rise to the concept of fuzzy  $\delta_i$ -complex feature. A combination of values  $a=(a_{i1},\dots,a_{ip})$  for the features of a support set  $\Omega = \{x_{i_1}, \dots, x_{i_p}\}$ , forms a  $\delta_i$ -complex fuzzy feature  $(a,\Omega)$  for  $K_i$ , with degree  $\mu_i((a,\Omega))$ ,  $i=1,\dots,r$ , if and only if: a)  $\exists O \in K'_i \mid x_{i1}(O) = a_{i1}$

$\wedge \dots \wedge x_{ip}(O) = a_{ip}$ ; b)  $\sum_{j=1}^m \Gamma(\Omega O_j, a) \mu_i(O_j) \geq \delta_i$ ; c)  $\sum_{j=1}^m \Gamma(\Omega O_j, a) (1 - \mu_i(O_j)) \leq \delta'_i$ ; and d)

$$\mu_i((a,\Omega)) = \frac{\sum_{j=1}^m \Gamma(\Omega O_j, a) \mu_i(O_j)}{\sum_{j=1}^m \mu_i(O_j)}, \text{ where } \delta_i > 0 \text{ and } \delta'_i > 0 \text{ are thresholds. When working with}$$

crisp classes, the assumption is that  $\mu_i((a,\Omega))=1$  for all  $(a,\Omega)$  associated with  $K_i$   $i=1,\dots,r$ .

This redefinition allows (as formerly mentioned) to handle any system of support sets. In

addition, it is considered in the expression the case of classes being fuzzy and each  $\delta_i$ -

complex feature has associated a degree of membership to said set, calculated in the basis

of the similarity of the combination of values  $a$  with the corresponding subdescriptions in

the respective class. The set of all  $\delta_i$ -complex fuzzy sets for  $K_i$  is denoted as  $RC(K_i)$ . On the

other hand, the set of all objects  $O \in K_i$  such that  $\sum_{(a,\Omega) \in RC(K_i)} \Gamma(\Omega O, a) < \eta_i$  will be called  $\eta_i$ -

fuzzy reminder of the class  $K_i$  and is denoted by  $r(K_i)$ . In the same manner as in the Kora-3

algorithm, the complementary complex features are computed in the remainder of the classes, but now using the new formulation. To each complex feature and each complementary complex feature, one computes a weight or importance  $P((a, \Omega))$  as

$$P((a, \Omega)) = \sum_{x_k \in \Omega} P(x_k) \sum_{O_j \in K_i'} \Gamma(\Omega O, a) P(O_j)$$

where  $P(x_k)$  and  $P(O_j)$  are the weights of feature  $x_k$  and of object  $O_j$ . Finally, to classify a new object  $O$ , it is compared with all  $\delta_i$ -complex fuzzy features of each class  $K_i$   $i=1, \dots, r.$ , in the following manner:

$$\mu_i(O) = \max_{(a, \Omega) \in RC(K_i')} \frac{\sum_{(a, \Omega) \in RC(K_i')} \Gamma(\Omega O, a) \mu_i((a, \Omega)) P((a, \Omega))}{\sum_{(a, \Omega) \in RC(K_i')} \Gamma(\Omega O, a) P((a, \Omega))}.$$

### Model based on representative sets

This model [42] is based in partial precedences and representative sets. The idea is to evaluate information in favor and against objects belonging to classes, and to consider that the parameters used for classification should be associated with each class. In contrast with previous models, this model can use a different system of support sets for each class. The rationale is that, for a particular class, a combination of features or ranges in values can provide valuable information to characterize such class, but one can not conclude that the same combination behaves similarly for another class. Hence, the model proposes to determine the system of support sets for each class.  $\{\Omega\}_j$  denotes the system of support sets associated with class  $K_j$  and  $CK_j$  represents the complement of the class. We define the *positive representatives set*  $M_j^1$  for class  $K_j$  with respect to  $\Omega \in \{\Omega\}_j$  as the set of all values corresponding to  $\Omega$  in the subdescriptions of the objects in  $K_j$  that are present  $\eta_j$  times and are not present in  $CK_j$ . Similarly, the *negative representatives set*  $M_j^0$  for class  $K_j$  with

respect to  $\Omega$  is defined as the set of all values corresponding to  $\Omega$  in the subdescriptions of the objects in  $CK_j$  that are present  $\eta_j$  times and are not present in  $K_j$ . The set of all values corresponding to  $\Omega$  in the subdescriptions of the objects of  $K_j$  that are not elements neither of the negative representatives set nor of the positive representatives set form the *neutral set of combinations* for class  $K_j$  denoted by  $M_j^\Delta$ . The algorithm rests on these three concepts. This method differentiates between the informational value of features and of objects, which can be calculated in some other way, see [20, 25], or be given by the expert. The first proposal of this model is Boolean, that is, considers descriptions of objects formed by Boolean features only. Thus, the algorithm is: 1) define the systems of support sets (and their associated parameters) for each class; 2) compute the sets  $M_j^1$ ,  $M_j^0$  and  $M_j^\Delta$  for each class; 3) to classify a new object  $O$ , compute  $\varphi_j(\Omega O)$  for each support set  $\Omega$  of each class  $K_j \quad j=1, \dots, r$ . This expression is defined in three cases: a)  $\varphi_j(\Omega O) = (p(x_{i_1}) + \dots + p(x_{i_s})) (p(\Omega O_{i_1}) + \dots + p(\Omega O_{i_s}))$  if  $\Omega O \in M_j^1$ , where  $p(x_i)$  is the weight of the variables conforming  $\Omega$  and  $p(\Omega O_i)$  means the weight of object  $O_i \in K_j$  that coincides with  $O$  according to  $\Omega$ ;

b)  $\varphi_j(\Omega O) = (p(x_{i_1}) + \dots + p(x_{i_s})) (p(\Omega O_{i_1}) + \dots + p(\Omega O_{i_s}))$  if  $\Omega O \in M_j^0$ , where  $p(x_i)$  is as before, and  $p(\Omega O_i)$  means the weight of object  $O_i \in CK_j$  that coincides with  $O$  according to  $\Omega$ ;

c)  $\varphi_j(\Omega O) = 0$  if  $\Omega O \in M_j^\Delta$ . 4) In this step the total evaluation  $\varphi_j(O)$  is computed for each class  $K_j \quad j=1, \dots, r$ , as  $\varphi_j(O) = \frac{1}{|\{\Omega\}_j|} \sum_{\Omega \in \{\Omega\}_j} \varphi_j(\Omega O)$ . 5) Finally, compute the r-tuple  $(\alpha_1(O), \dots, \alpha_r(O))$  in which one finds the membership relation  $\alpha_j(O)$  of object  $O$  for each one

of the  $r$  classes, as follows:  $\alpha_j(O) = \begin{cases} 1 & \text{if } \varphi_j(O) > 0 \\ 0 & \text{if } \varphi_j(O) < 0 \\ * & \text{if } \varphi_j(O) = 0 \end{cases}$ . This model proved to be of limited

use for solution of practical problems, where objects are described by features that can take any kind of values. Thus, an improvement is proposed in [25], which allows such descriptions (and values) but defines the comparison criteria to yield Boolean values. In this manner, a similarity function among two subdescriptions is defined, which can be used to determine if a subdescription belongs to  $M_j^1$ ,  $M_j^0$  or  $M_j^\Delta$  in the same manner as in the Boolean case, but now verifying if they are similar or not. For this, the following expressions are evaluated:

$$\varphi_j^+(\Omega O) = \left( p(x_{i_1}) + \dots + p(x_{i_s}) \right) \left[ \chi_1 \sum_{O_i \in K_j} p(O_i) \Gamma(\Omega O, \Omega O_i) + \chi_0 \sum_{O_i \in CK_j} p(O_i) \bar{\Gamma}(\Omega O, \Omega O_i) \right] \quad \text{and}$$

$$\varphi_j^-(\Omega O) = \left( p(x_{i_1}) + \dots + p(x_{i_s}) \right) \left[ \chi_1 \sum_{O_i \in CK_j} p(O_i) \Gamma(\Omega O, \Omega O_i) + \chi_0 \sum_{O_i \in K_j} p(O_i) \bar{\Gamma}(\Omega O, \Omega O_i) \right], \quad \text{where}$$

$p(x_i)$ ,  $p(O_i)$  are weights associated to the features and objects,  $\chi_0$ ,  $\chi_1$  are weight factors and

$\bar{\Gamma}$  takes values opposite to those of  $\Gamma$ . We then decide if  $\Omega O$  is in  $M_j^1$ ,  $M_j^0$  or  $M_j^\Delta$  as

follows: a)  $\Omega O \in M_j^1$  if  $\varphi_j^+(\Omega O) - \varphi_j^-(\Omega O) > \delta$ ; b)  $\Omega O \in M_j^0$  if  $\varphi_j^-(\Omega O) - \varphi_j^+(\Omega O) > \delta$ ; and

c)  $\Omega O \in M_j^\Delta$  if  $|\varphi_j^+(\Omega O) - \varphi_j^-(\Omega O)| < \delta$ . Except for these changes, the previous algorithm

remains the same.

Carrasco [43] proposes an extension to [42], in which objects can be described by features of any kind, classes do not need to be disjoint; they can be fuzzy, although imposing the restriction that for every class there must exist at least one object that is closer to totally

belonging to the class than of not belonging to the class, which is reasonable, since this is a problem of supervised classification. This extension is reached by extending the concept of positive, negative and neutral representatives sets, as follows. Let  $\Omega$  be a support set for class  $K_j$ . The positive representatives set for class  $K_j$  with respect to  $\Omega$ , denoted by  $M_j^1$ , is

defined as the set of all  $\Omega O_p$  such that: 1)  $\exists O \in MA \mid \Omega O = \Omega O_p$ ;

$$2) \sum_{i=1}^m \Gamma(\Omega O_i, \Omega O_p) \alpha_j(O_i) \geq \eta_j; \quad 3) \sum_{i=1}^m \Gamma(\Omega O_i, \Omega O_p) (1 - \alpha_j(O_i)) < \delta_j;$$

$$4) \mu_j(\Omega O_p) = \frac{\sum_{i=1}^m \Gamma(\Omega O_i, \Omega O_p) \alpha_j(O_i)}{\sum_{i=1}^m \alpha_j(O_i)}, \text{ where } \mu_j(\Omega O_p) \text{ denotes the degree by which the}$$

combination of values  $\Omega O_p$  is positive representative of  $K_j$ ;  $\alpha_j(O)$  is the degree of membership of  $O$  to the class  $K_j$  and  $(1 - \alpha_j(O))$  represents the degree of membership to the complement of  $K_j$ ,  $\eta_j$  represents the minimum similarity threshold of  $\Omega O_p$  with all the objects in  $K_j$ ;  $\delta_j$  represents the maximum similarity threshold of  $\Omega O_p$  with respect to all objects in the complement of  $K_j$ . Analogously, the negative representatives set  $M_j^0$  is

redefined by interchanging in expressions 2), 3) and 4),  $\alpha_j(O)$  by  $(1 - \alpha_j(O))$ . The

combinations  $\Omega O_p$  that are not in  $M_j^1$  nor in  $M_j^0$  will be neutral combinations and shall be denoted by  $M_j^\Delta$ . The informational relevance of a positive (negative) representative  $\Omega O_p$

for a class  $K_j$ , is denoted by  $P(\Omega O_p)$ , and it is computed by the expression

$$P(\Omega O_p) = (p(x_{i_1}) + \dots + p(x_{i_s})) \sum_{i=1}^m p(O_i) \alpha_j(O_i) \Gamma(\Omega a, \Omega O_i). \text{ Given an object } O \text{ to classify,}$$

the total positive evidence, denoted by  $\varphi_i^+(O)$ , and the total negative evidence, denoted by

$\varphi_i^-(O)$ , are computed for each one of the classes  $K_j, j=1, \dots, r$  as follows:

$$\varphi_i^+(O) = \begin{cases} \sum_{\Omega \in \{\Omega_1^i, \dots, \Omega_{s_i}^i\}} \sum_{\Omega O_p \in \Omega M_i^1} \Gamma(\Omega O, \Omega O_p) P(\Omega O_p) & \text{if the classes are crisp sets} \\ \sum_{\Omega \in \{\Omega_1^i, \dots, \Omega_{s_i}^i\}} \sum_{\Omega O_p \in \Omega M_i^1} \Gamma(\Omega O, \Omega O_p) P(\Omega O_p) \mu(\Omega O_p) & \text{if the classes are fuzzy sets} \end{cases}$$

$$\varphi_i^-(O) = \begin{cases} - \sum_{\Omega \in \{\Omega_1^i, \dots, \Omega_{s_i}^i\}} \sum_{\Omega O_p \in \Omega M_i^0} \Gamma(\Omega O, \Omega O_p) P(\Omega O_p) & \text{if the classes are crisp sets} \\ - \sum_{\Omega \in \{\Omega_1^i, \dots, \Omega_{s_i}^i\}} \sum_{\Omega O_p \in \Omega M_i^0} \Gamma(\Omega O, \Omega O_p) P(\Omega O_p) \mu(\Omega O_p) & \text{if the classes are fuzzy sets} \end{cases}$$

and then the total evaluation  $\varphi_j(O)$  is computed for each class  $K_j, j=1, \dots, r$ , as follows:

$$\varphi_j(O) = \begin{cases} 0 & \text{if } \varphi_j^+(O) = \varphi_j^-(O) = 0 \\ \frac{\varphi_j^+(O) + \varphi_j^-(O)}{2 \max\{\varphi_j^+(O), |\varphi_j^-(O)|\}} & \text{otherwise} \end{cases}$$

This evaluation estimates the membership of  $O$  in class  $K_j$ . 1) If working with crisp classes,

$$\text{use: } \alpha_j(O) = \begin{cases} 1 & \text{if } \varphi_j(O) > 0 \\ 0 & \text{if } \varphi_j(O) < 0 \text{ in this case * denotes abstention.} \\ * & \text{if } \varphi_j(O) = 0 \end{cases}$$

2) For fuzzy classes use:

$$\alpha_j(O) = \begin{cases} 0.5 + \varphi_j(O) (2 \max_{O_i \in MA} \{\mu_j(O_i)\} - 1) & \text{if } \varphi_j(O) \geq 0 \\ 0.5 + \varphi_j(O) & \text{if } \varphi_j(O) < 0 \end{cases}$$

A detailed study of supervised classification algorithms based in partial fuzzy precedence is given in [44].

In Logical Combinatorial Pattern Recognition, some other models of supervised classification have been proposed. Such is the case of algorithms based in the weight of the patterns [25, 45], among which we find: voting algorithms; those based in the ideal of the class, and algorithms based in exactness thresholds.

Another model of supervised classification is the algorithms based in typicality and contrast. This model was initially proposed by Diordenko and Vaskovskii [46] for the solution of problems with geologic anomalies, its purpose was not to classify geological regions, but to structure their data so as to facilitate the classification. Based on these ideas, Vega [47] extended the algorithm and developed a supervised classification algorithm based in notions of typicality and contrast. This particular algorithm has been used to classify patients with cleft palate [48].

## **UNSUPERVISED CLASSIFICATION**

In this problem it is assumed that the structure of universe  $U$  is not known. To find such structure, in general a subset  $MI$  of  $U$ , the initial sample, is given. On occasions the number of groups that there should result (or the number desired), is known, and the problem is termed *restricted unsupervised classification*. On other occasions, this number is also unknown (*free unsupervised classification*). Here, only a representation space that is a Cartesian product without further properties is used. The problem is precisely to find the classes, the groupings. Such groups need not be formed by crisp sets.

### **Free unsupervised classification**

The first tools in this field appear in Sirotinskaia [49], who proposes the CLASS algorithm. This is a hierarchical algorithm and was written to solve geology prospecting problems. It is based in the concept of a compact set. A set  $NU_r$  is a compact set if and only if

$$a) \forall O_j \in MI$$

$$[O_i \in NU_r \wedge (\max_{\substack{O_t \in MI \\ O_t \neq O_i}} \{\Gamma(O_i, O_t)\} = \Gamma(O_i, O_j) \geq \beta_0 \vee \max_{\substack{O_t \in MI \\ O_t \neq O_j}} \{\Gamma(O_j, O_t)\} = \Gamma(O_j, O_i) \geq \beta_0)] \Rightarrow O_j \in NU_r;$$

$$b) \forall O_i, O_j \in NU_r \exists O_{i_1}, \dots, O_{i_q} \in NU_r$$

$$[O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p \in \{1, \dots, q-1\}; [\max_{\substack{O_t \in MI \\ O_t \neq O_{i_p}}} \{\Gamma(O_{i_p}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0 \vee$$

$$\max_{\substack{O_t \in MI \\ O_t \neq O_{i_{p+1}}}} \{\Gamma(O_{i_{p+1}}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0]]$$

c) Every isolated object constitutes a  $\beta_0$ -compact (degenerate) set. This definition was proposed by Martínez [50,51]. The definition of [49] does not appear here.

The CLASS algorithm is summarized as follows: step 1) Compute the values of all possible pairs of MI; that is, form a similarity matrix  $MS = \|\Gamma(O_i, O_j)\|_{m \times m}$  where  $m$  is the number of

objects of MI; step 2) Compute  $\beta_0 = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \Gamma(O_i, O_j)$ ; step 3) Compute the

$\beta_0$ -compact sets; step 4) Compute the similarity measures among all compact sets

$NU_1, \dots, NU_c$  using the expression:  $\Gamma(NU_s, NU_v) = \frac{1}{|NU_s| |NU_v|} \sum_{t=1}^{|NU_s|} \sum_q^{|NU_v|} \Gamma(O_t, O_q)$ ; step

5) compute a new  $\beta_0'$  from the matrix of comparisons among nuclei, and repeat steps 4 and

5. In this same sense Ruiz *et al.* [29] propose a modification based in the HOLOTYPE algorithm (originally proposed by Y.A. Voronin *et al.* [52] for supervised classification) giving rise to a new algorithm analogous to CLASS but now using the concept of

connected set –which is the one used by HOLOTYPE. A set  $NU_r$  is a *connected set* if and only if: a)  $\forall O_i, O_j \in NU_r \exists O_{i_1}, \dots, O_{i_q} \in NU_r$

$$[O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p \in \{1, \dots, q-1\} \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0]$$

$$b) \forall O_i \in MI [(O_j \in NU_r \wedge \Gamma(O_i, O_j) \geq \beta_0) \Rightarrow O_i \in NU_r]$$

c) Every  $\beta_0$ -isolated object is a  $\beta_0$ -*connected* (degenerate) set. This definition was proposed by Montellano [53, 54]. The definition of connected set of [52] does not appear here.

Also, the TAXDIF model [53, 54], based on graphs, which allows the descriptions of objects to be given by features of any kind, was developed. It proposes crisp and fuzzy clustering criteria for grouping of objects, similar to the compact and connected criteria used in CLASS and HOLOTYPE. A significant contribution in this new model is the definition of crisp grouping criterion  $\Pi$ . This definition generates several grouping criteria, including the connected and compact criteria. Similarly, the model defines the concept of fuzzy grouping criterion  $\Pi_d$  which allows, from a crisp structuring generated by a crisp criterion  $\Pi$ , the generation of a fuzzy structuring of the original sample. That is, in the result, each object belongs in different degree to each of the clusters (clouds). Also, TAXDIF is flexible in handling several similarity functions, which can be partial functions and be defined using support sets, such as those used in supervised classification, see [25].

As example, the membership function defined for connected sets is:

$$\mu_{NB_r}(O_i) = \begin{cases} 1 & \text{if } NU_r = \{O_i\} \\ \max_{O_j \in NU_r \setminus \{O_i\}} \{\Gamma(O_i, O_j)\} & \text{otherwise} \end{cases} . \text{ The detailed definition of other fuzzy}$$

criteria is given in [51]. Membership of an object in a crisp connected set is given by the

fulfillment of certain similarity properties, which establish the definition of crisp grouping criterion. Similarly, the degree of membership of an object  $O_i$  to a connected (fuzzy) cloud is given in terms of the degree of fulfillment of said properties.

This model is an alternative to the C-means fuzzy algorithm, which is defined for numeric features. The model of TAXDIF algorithms can be described by the following steps: step 1) Define the similarity function  $\Gamma$ ; step 2) Determine threshold  $\beta_0$ ; step 3) Define the fuzzy grouping criterion  $\Pi_d$ ; step 4) Generate the family of crisp groupings  $\tau$ ; step 5) Define the family of fuzzy groupings or clouds corresponding to  $\tau$ . Varying the parameters results in different algorithms.

Due to the features of the model and its potential in the solution of problems such as those present in soft-sciences, a theoretical study was performed, about the set relations among the groupings generated by several grouping criteria.

Fig. 1

This study shows that for the same  $\beta_0$  value, the structures (groupings) found for several grouping criteria are related (see figure 1). For instance, connected groupings are unions of compact groupings. Other relations are in Martínez *et al.* [51]. This hierarchy imposes an order among the structures or groupings, following their generality, which can be of use in practice. Low levels contain very specific categories (including isolated prominent objects); upper levels are formed by more general structures. This disposition of said structures can provide, for a same universe of objects, different visions, with different abstraction levels. [55] proposes these tools for the solution of semantic grouping problems, in this case

objects are related key words from documents, and the problem is to find word groupings based in semantic similarity. [27] proposes an algorithm for constructing classes of a thesaurus.

Now a days, unsupervised classification tools for large information volumes are being developed [56]. Also, conceptual unsupervised clustering has been proposed [57-60]. In these works one finds contributions to the so-called (in machine learning) unsupervised learning or conceptual clustering. The problem is: from a set of objects, find how these objects are grouped and, in addition, find the properties or concepts that objects in each group satisfy. This research line has aroused much interest, in recent years, for its potential in data mining problems.

### **Restricted unsupervised classification**

Logical Combinatorial Pattern Recognition has devoted less effort to these problems. Nevertheless, algorithms have been proposed where the number of desired clusters is specified *a priori*. For instance, García and Martínez [61, 62] proposed a modification to the C-means algorithm allowing object descriptions to be of any nature. Keeping the original structure of the algorithm, they use, instead of a distance measure, a similarity measure among objects, such as those of [25]. The similarity function must fulfil: 1)  $\Gamma(O_j, O_k) \in [0,1]$  for  $1 \leq j \leq m$  and  $1 \leq k \leq m$  ( $m$  is the number of objects to group); 2)  $\Gamma(O_j, O_j) = 1$  for  $1 \leq j \leq m$ ; and 3)  $\Gamma(O_j, O_k) = \Gamma(O_k, O_j)$  for  $1 \leq j \leq m$  and  $1 \leq k \leq m$ . Let  $u_{ik}$  be the membership degree of object  $O_k$  in grouping  $C_i$ , and let  $R^{c \times m}$  be the set of all real matrices  $c \times m$ . Any  $c$ -partition [61] of the data set is represented by a matrix  $U = [u_{ik}] \in R^{c \times m}$ , which satisfies: 1)  $u_{ik} \in \{0,1\}$  for  $1 \leq j \leq m$  and  $1 \leq k \leq m$ ; 2)  $\sum_{i=1}^c u_{ik} = 1$  for  $1 \leq k \leq m$ ; and 3)  $\sum_{k=1}^m u_{ik} > 0$  for

$1 \leq i \leq c$ . Then, the partition matrix  $U$  is determined by the maximization of the objective function given by  $J(U) = \sum_{i=1}^c \sum_{k=1}^m u_{ik} \Gamma(O_i^r, O_k)$  where  $\Gamma(O_i^r, O_k)$  is the similarity between the most representative object  $O_i^r$  (“the center”) in grouping  $C_i$  and object  $O_k$ . Notice that in this case “the center” is an object belonging to the sample, instead of a fictitious element such as in classical C-means. Working with a similarity measure (and not with a distance) implies that the calculation from seeds can not be realized by using centroids (as the original work does, since it works in multidimensional metric spaces). [61, 62] proposes a way to compute these representative objects, using similarity. Once these elements are found the functional  $J(U)$  is maximized when  $u_{ik}$  is determined as follows:

$$u_{ik} = \begin{cases} 1 & \text{if } \Gamma(O_i^r, O_k) = \max_{1 \leq q \leq c} \{\Gamma(O_q^r, O_k)\} \\ 0 & \text{otherwise} \end{cases} . \text{ That is, object } O_k \text{ shall be assigned to the cluster}$$

having its representative object most similar with  $O_k$ .

Thus, the algorithm is: step 1) Select  $c$  objects in the sample as initial seeds. Fix  $ni'$ , the maximum number of iterations, and set  $ni=0$ ; step 2) Compute the partition matrix  $U=U^{(ni)}$  using the previously mentioned initial seeds; step 3) Determine the representative objects of the groupings of matrix  $U^{(ni)}$ ; step 4) Compute the partition matrix  $U^{(ni+1)}$ ; step 5) If the set of representative objects is the same as in the last iteration, stop. Otherwise, increment  $ni=ni+1$ ; step 6) If  $ni > ni'$  stop. Else, go to step 3.

More recently, other restricted unsupervised classification algorithm [64] has been proposed, based in the concept of connected set. By introducing suitable concepts and results, this algorithm works with any similarity function and with data described by features of any kind. Also, it determines theoretically how many different ways there are to

partition a space into connected components, the value of thresholds  $\beta_0$  characterizing them, and the cardinality of each of these partitions.

### **Acknowledgments**

One of the authors (A. G.) acknowledges the hospitality of Dr. Robert S. Ledley, Editor of *Pattern Recognition*, for the opportunity to be Guest Editor of the Special Section on non conventional pattern classifiers, in this issue.

### **REFERENCES**

IPN stands for Instituto Politécnico Nacional, Mexico City.

CINVESTAV stands for Centro de Investigación y Estudios Avanzados del IPN.

CIC stands for Centro de Investigación en Computación del IPN.

*I* through *V* *TIARP* (also called *SIARP*) stand for *Iberoamerican Workshop on Pattern Recognition*, as follows: *I TIARP*, Havana (1995); *II TIARP*, Havana (1997); *III TIARP*, CIC-IPN, Mexico City (1998); *IV SIARP*, Havana (1999); *TIARP2000*, Acapulco (2000, in conjunction with *Mexican International Conference on A. I. MICAI 2000*) and *V SIARP* Lisboa (2000). Proceedings are available from CIC-IPN.

1. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, London (1990).
2. H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psych.* **24**, 417-441, 498-520 (1933).
3. L. L. Thrustone, Multiple Factor analysis, *Psych. Rev.* **38**, 406-427 (1931).

4. Robert J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches*, John Wiley & Sons, Inc. (1992).
5. J.T. Tou and R. Gonzalez, *Pattern Recognition Principles*. Addison Wesley, Reading, Ma. (1974).
6. R. O. Duda and P. E. Hart, *Pattern classification and scene analysis*. John Wiley & Sons (1973).
7. G. Ball and D. J. Hall, ISODATA, An iterative method of multivariate data analysis and pattern classification. *Proc. IEEE International Communications Conference*. Philadelphia (1966).
8. G. Ball and D. Hall, A clustering technique for summarizing multivariate data. *Behav. Sci.* **2**, 153-155 (1967).
9. K. Jain, R.C. Dubes, *Algorithms for clustering data*, Prentice Hall (1998).
10. J. G. Augustson and J. Minker, An analysis of some graph theoretical clustering techniques, *J. ACM* **17**, 571-588 (1970).
11. C. T. Zahn, Graph theoretical methods for detecting and describing gestalt clusters, *IEEE Trans. Comp. C-20*, 68-86 (1971).
12. M. Lazo-Cortés, J. Ruiz-Shulcloper, and E. Alba-Cabrera. An overview of the evolution of the concept of testor in pattern recognition. *Pattern Recognition* (this issue).
13. A.N. Dmitriev, Yu. I. Zhuravlev, and F.P Krendelev, About the mathematical principles of objects and phenomens classification, *Sbornik Diskrtnii Analisis* **7**, 3-15. Novosibirsk (1966). In Russian.
14. M. Ortiz-Posadas, *Elaboration and application of a new pattern recognition methodology in Medicine*, Ph. D. Thesis, Universidad Autónoma Metropolitana, Iztapalapa Campus, Mexico City (1999). In Spanish.

15. J. Ruiz-Shulcloper, E. Alba-Cabrera, and M. Lazo-Cortés, *Introduction to pattern recognition: logical-combinatorial method*, Technical Report No. 51, Green Series, CINVESTAV (1995). In Spanish.
16. H. A. Heathfield, G. Winstanley, and N. Kirkham, Decision support system for the differential diagnosis of breast disease. *J. Biomed. Eng.* **13**, 51-57 (1991).
17. M. Ortiz-Posadas, F. Martínez-Trinidad, and J. Ruiz-Shulcloper, A new approach to differential diagnosis of diseases, *Int. J. Biomed. Comput.* **40**, 179-185 (1996).
18. M. R. Ortiz-Posadas, Prognosis and evaluation of cleft palate patients rehabilitation using pattern recognition techniques, *World congress on medical physics and Biomedical Engineering* **35**, 1, 500. Niza, France (1997).
19. J. Ruiz-Shulcloper, E. Alba-Cabrera, N. López-Reyes, M. Lazo-Cortés, and E. Barreto-Fiu, *Topics about Testor Theory*. Technical report, Yellow series, No. 134, CINVESTAV (1994). In Spanish.
20. M. Lazo-Cortés and Ruiz-Shulcloper, Determining the feature relevance for non-classically described objects and a new algorithm to compute typical fuzzy testors, *Pattern Recognition Letters* **16**, 1259-1265 (1995).
21. J. Carrasco-Ochoa, *Sensibility in Logical-combinatorial Pattern Recognition*, Ph. D. Thesis in preparation, CIC-IPN (2000). In Spanish.
22. Y. I. Zhuravlev, About the algebraic method for solution of recognition or classification problems, *Problemi kibernetiki* **33**, 5-68 (1978). In Russian.
23. V. Valev and Y. I. Zhuravlev, Integer-valued problems of transforming the training tables in k-valued code in pattern recognition problems, *Journal of Pattern Recognition* **24**, 4, 283-288 (1991).

24. Goldman, R. S, Problems of the theory of typical testors, *Avtomatika i telemekhanika* **10**, 143-153 (1980). In Russian.
25. J. Ruiz-Shulcloper, A. Guzman-Arenas, and J. F. Martínez-Trinidad, *Logical Combinatorial Approach to Pattern Recognition: I. Feature selection and supervised classification*. Editorial Politecnica, Mexico (1999). Also: Fondo de Cultura Economica, Mexico (2000). In Spanish.
26. J. Ozols and A. Borisov, Fuzzy classification based on patterns projections analysis, *Pattern Recognition* (this issue).
27. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, Fuzzy clustering of semantic spaces, *Pattern Recognition* (this issue).
28. Y. I. Zhuravlev, M. M. Camilov, and Sh. E. Tulianganov, Algorithms for computing the evaluations and their applications. Tashkent FAN (1974). In Russian.
29. J. Ruiz-Shulcloper and M. Lazo-Cortes, *Mathematical Models for Pattern Recognition*. Ed. UCLV, Santa Clara, Cuba (1990). In Spanish.
30. J. Gómez-Herrera *et al.* Prognostic of Gas-oil deposits in the Cuban ophiological association, applying mathematical modeling. *Geophysics International* **33**, 3, 447-467 (1995). In Spanish.
31. S. López-Pérez, M. Lazo-Cortes, and H. M. Estrada-García, Medical Electro-diagnostic using pattern recognition tools, *Proc. II TIARP* 237-244. In Spanish.
32. J. Ruiz-Shulcloper *et al.*, Mathematical modeling and prognostic of maximum magnitude of earthquake in the Caribbean region. Recognition of Spatial structures, *Academia* 81-101, Cuba (1992). In Spanish.
33. J. Ruiz-Shulcloper, PROGNOSIS and its applications to Geo-sciences, *Proc. Third Iberoamerican Congress on Artificial Intelligence* 561-586, Havana (1992). In Spanish.

34. M. Lazo-Cortés, Models based on Testor Theory for feature selection and supervised classification with non-classically object descriptions, Ph. D. Thesis, Central University from Las Villas, Santa Clara, Cuba (1994). In Spanish.
35. L. A. Zadeh, Fuzzy Sets, *Information and Control* **8**, 338-353 (1965).
36. H. J. Zimmermann, *Fuzzy set theory and its applications*, Kluwer Academic Publishers. Boston (1991).
37. G.J. Klir, T.A. Folger, *Fuzzy sets, uncertainty and information*, Prentice Hall. New Jersey (1988).
38. M. N. Bongard *et al.*, Solution to geological problems with support of recognition programs, *Sov. Geologia*, **6** (1963).
39. V. Keilis-Borok, A. Soloviov, Pattern Recognition in Earthquake prediction, *Workshop in Non-linear Dynamics and Earthquake prediction* 1-14. International Center for Science and High Technology, Trieste, Italy (1991).
40. V. Keilis-Borok and A. Soloviov. Pattern Recognition: General Description. *Workshop in Non-linear Dynamics and Earthquake prediction* 1-14. International Center for Science and High Technology, Trieste, Italy (1991).
41. L. A. De la Vega-Doria, J. A. Carrasco Ochoa, and J. Ruiz Shulcloper, Fuzzy Kora- $\Omega$  Algorithm, *Proc. 6<sup>th</sup> European Congress on Intelligent Techniques and Soft Computing* 1190-1194. Aachen, Germany (1998).
42. L. Baskakova and Y. I. Zhuravliev, Model of recognition algorithms with representative sets and systems of support sets, *Zh. Vichislitelnoi Matematiki i Matematicheskoi Fiziki* **2**, 5, 1264-1275 (1981). In Russian.

43. J. A. Carrasco-Ochoa and J. Ruiz-Shulcloper, A model of supervised classification based in representative sets, *Proc. III TIARP* 267-278. In Spanish.
44. J. Ruiz-Shulcloper and M. Lazo-Cortés, Mathematical Algorithms for the Supervised Classification Based on Fuzzy Partial Precedence, *Mathematical and Computer Modelling* **29**, 4, 111-119 (1999).
45. A. Fuentes, J. Ruiz-Shulcloper, and L. Romero, Classification Algorithms based on the patterns' weights, *Ciencias Matemáticas* **IV(1)**, 97-121 (1983). In Spanish.
46. L. Diordenko and B. Vaskovskii, Classification algorithm to determine AGE anomalies, *Problemy Kibernetiki* No. 23, 131-145.
47. L. Vega-Alvarado, An algorithm for data analysis and classification type typicality and contrast, M. Sc. Thesis. CINVESTAV (1994). In Spanish.
48. L. Vega-Alvarado and M. R. Ortiz-Posadas, Analysis and classification of cleft palate patients through typicality and contrast concepts, *Proc. IV SIARP* 419-424. In Spanish.
49. S.V. Sirotinskaia, Logical methods for analysis of geological information, Ed. Nedra (1986). In Russian.
50. J. F. Martínez-Trinidad, System for data analysis and the structuralization of universes. M. Sc. Thesis, Universidad Autónoma de Puebla, Mexico (1995). In Spanish.
51. J. F. Martínez-Trinidad, J. Ruiz-Shulcloper, and M Lazo-Cortés. Structuralization of universes. *Fuzzy Sets & Systems* **112**, 3, 485-500 (2000).
52. Y. A. Voronin, G. N. Karataeva, E. N. Epshtein, *et al.*, *HOLOTYPE programs for solving problems of pattern recognition*, Ed. Alma Atá (1968). In Russian.
53. J. J. Montellano Ballesteros, Clusters in fuzzy graphs, M. Sc. Thesis. Universidad Nacional Autónoma de México, Mexico City (1994). In Spanish.

54. J. Ruiz-Shulcloper and J. J. Montellano-Ballesteros, A new model of fuzzy clustering algorithms *Proc. 3th European Congr. on Fuzzy and Intelligent Technologies* 1484-1488, Aachen, Germany (1995).
55. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, Fuzzy semantic clustering, *Proc. Fourth European Congress on Intelligent Techniques and Soft Computing* 1397-1401, Aachen, Germany (1996).
56. G. Sanchez-Díaz and J. Ruiz-Shulcloper, MID Mining: a Logical Combinatorial Pattern. Recognition approach to Clustering in Large Data Sets. Submitted to *15<sup>th</sup>. International Conference on Pattern Recognition*, Barcelona, Spain (2000).
57. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, Fuzzy Conceptual Clustering, *Proc 5th European Congr. on Fuzzy and Intelligent Technologies* 1852-1857, Aachen, Germany (1997).
58. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, Fuzzy LC conceptual algorithm. *Proc. 6th European Congr. on Fuzzy and Intelligent Technologies* 20-24, Aachen, Germany (1998).
59. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, Crisp LC-conceptual algorithm, *Proc. IV SIARP* 195-206. In Spanish.
60. J. F. Martínez-Trinidad and J. Ruiz-Shulcloper, LC-conceptual algorithm: Characterization using typical testors by class, *Proc. 7th European Congr. on Fuzzy and Intelligent Technologies* (On CD). Aachen, Germany (1999).
61. J. R. García-Serrano and J. F. Martínez-Trinidad, Extension to C-means algorithm for the use of similarity functions, *Proc. 3<sup>rd</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases* 354-359, Prague, (1999).

62. J. F. Martínez-Trinidad and J. R. García-Serrano, A new C-means algorithm using similarity functions, Submitted to *Int. Journal of Mathematical and Modelling Computing* (1999).
63. E. R. Ruspini, A new approach to clustering, *Information and control* No.15, 22-32 (1969).
64. R. Reyes-Gonzalez and J. Ruiz-Shulcloper, An algorithm for restricted structuralization of spaces, *Proc. IV SIARP* 267-278. In Spanish.