PERGAMON Mathematical and Computer Modelling 29 (1999) 111-119

MATHEMATICAL AND COMPUTER MODELLING

Mathematical Algorithms for the Supervised Classification Based on Fuzzy Partial Precedence

J. RUIZ-SHULCLOPER*
Instituto de Cibernética, Matemática y Física,
E # 309 esq. 15 Habana 10400, Cuba
and
Centro de Investigación en Computación-IPN
Juan de Dios Batiz s/n esq. Miguel Othón de Mendizabal,
Col. Nueva Industrial Vallejo (Lindavista),
C.P. 07738, México, D.F.
jruiz@pollux.cic.ipn.mx recpat@cidet.icmf.inf.cu

M. LAZO-CORTÉS**
Instituto de Cibernética, Matemática y Física,
E # 309 esq. 15 Habana 10400, Cuba

(Received June 1997; revised and accepted July 1997)

Abstract—Some mathematical algorithms for supervised classification problems are presented in this paper. These algorithms are based on fuzzy partial precedents, and they allow us to work with nonclassically described objects, i.e., mixed data. These types of descriptions frequently arise in soft sciences. As a rule, the most used methods for solving such classification problems are oriented towards one type of feature, most often quantitative. They do not allow the use of different kinds of features, as a consequence, all variables must be quantitative, or exclusively qualitative. In fact, those methods use, in some way, a distance measure between object descriptions, which follows from the hypothesis of compactness of classes. The proposed models allow the handling of quantitative and qualitative features together, and missing values. They are based on partial evaluation of similarity between objects in a fuzzy environment. They do not use a distance. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords—Fuzzy pattern recognition problems, Fuzzy supervised classification, Partial precedence.

INTRODUCTION

Supervised classification problems, on which objects are described by mixed data, are frequently found in soft sciences. In this case, specialists often consider the problem as a collection of subproblems, each one viewed from a different point of view. They analyze separately some sets of features converging to partial conclusions to finally make a decision. Some mathematical techniques for modeling this procedure are called "mathematical models based on partial precedents".

*Presently with Centro de Investigación en Computación, IPN México.
**Presently with Sección de Computación, CINVESTAV-IPN México.

The authors wish to thank the reviewers for their helpful suggestions and constructive criticism.

0895-7177/99/$ - see front matter © 1999 Elsevier Science Ltd. All rights reserved Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX
PII: S0895-7177(99)00044-8

A practical problem, solved in this way, can be found in [1]. In [2], the authors mention one of these models as: "combinatorial-logic method ... which construct informative fragments of the objects descriptions". In this paper, we propose some mathematical models, based on fuzzy partial precedents, to solve supervised classification problems in a fuzzy environment,

> Let $\mathbf{U}$ be a finite set of objects structured in $K_1, \dots, K_r$ classes. $r \geq 2$, described in terms of a finite set of features $\mathbf{R} = \{x_1, \dots, x_n\}$, each of which has associated a set of admissible values $\mathbf{M}_i$. Given $M = \{\mathcal{O}_1, \dots, \mathcal{O}_m\} \subseteq \mathbf{U}$, which is the union of $K'_1, \dots, K'_r, K'_i \subset K_i$, $i = 1, \dots, r$ and $\mathcal{O} \in (\mathbf{U} \backslash M)$, the membership relation between $\mathcal{O}$ and $K_1, \dots, K_r$ has to be found. (1)

In general terms, this is the supervised classification problem to be solved. The following set of assumptions is frequently implicit.

a) $\mathbf{M}_i \subseteq \mathbb{R}$, or exclusively, $\mathbf{M}_i = \{0, 1\}$ for every $i = 1, \dots, n$, which means that we are working on a subspace of $\mathbb{R}^n$, or of the n-dimensional Boolean space B^n.
b) A metric $\mathbf{d}$ is defined on $\mathbf{M}_1 \times \dots \times \mathbf{M}_n$.
c) $\mathbf{d}$ induces a metric on every $\mathbf{M}_i$, which means that comparisons between values of the same variable, as well as those between objects descriptions, are made regarding their "closeness", and performed by distances on $\mathbb{R}^n$ or B^n.
d) $K_1, \dots, K_r$, are crisp sets.
e) The classes are disjoint sets.
f) The similarity relations between objects of the universe are determined through comparisons of their complete descriptions.
g) The response of the classifiers is, at most, one class.
h) Some statistical behaviors are assumed on the value distribution of the variables, and missing values are not allowed.

To overcome this last situation, interpolation techniques with implicit restrictions regarding the nature and behavior of their value are applied.

In practice, it is easy to find problems in soft sciences on which most of these assumptions are not true, and therefore, many authors build their models following assumptions which we do not accept to be right (see, for example, [3]). Therefore, the development of correct mathematical tools which allow us to obtain a reasonable solution to the above-mentioned problems, without misrepresenting or forcing their true nature, is a goal worthy of many additional efforts. This paper concerns one effort in this direction.

FORMAL STATEMENT OF THE PROBLEM IN FUZZY ENVIRONMENT

Let us consider a supervised classification problem in the following terms.

Let $\mathbf{U}$ be the same as before. We do not make any assumptions regarding the set of variables $\mathbf{R}$, so they may be either qualitative, quantitative, fuzzy, or linguistic. The admissible sets of values $\mathbf{M}_i$ may also take the $*$ value to denote a missing value. For every $\mathcal{O}$ element of $\mathbf{U}$, there exists a point in $\mathbf{M}_1 \times \dots \times \mathbf{M}_n$, that we will denote $D(\mathcal{O}) = (x_1(\mathcal{O}), \dots, x_n(\mathcal{O}))$, where $x_i(\mathcal{O}) \in \mathbf{M}_i$ represents the evaluation of the variable x_i $(i = 1, \dots, n)$ for the object $\mathcal{O}$; $D(\mathcal{O})$ will be called object description.

Let us assume that classes $K_1, \dots, K_r$ are fuzzy subsets of $\mathbf{U}$. Every object $\mathcal{O} \subset \mathbf{U}$ may be associated with a membership r-tuple, $\alpha(\mathcal{O}) = (\alpha_1(\mathcal{O}), \dots, \alpha_r(\mathcal{O}))$ such that $\alpha_j(\mathcal{O}) = \mu_{k_j}(\mathcal{O})$ is the membership degree of the object $\mathcal{O}$ to the class K_j, $j = 1, \dots, r$.

Let φ_i be a function

$$\varphi_i : \mathbf{M}_i \times \mathbf{M}_i \longrightarrow V \tag{2}$$

which states the comparison between two values of the variable x_i. V will be $\{0, 1\}$ for instance, whenever the result of the comparison is exactly one of two possible answers (similar or not similar). V may be considered as $[0, 1]$ if the answer is, for instance, of a fuzzy nature (degrees of differentiation). If a metric is used as φ_i, then V will be $\mathbb{R}_+$. V may also be the set of terms of a linguistic variable.

Another problem that arises in practice is that specialists, very often, do not perform a comparative analysis between two objects, paying simultaneous attention to the complete set of the features. Instead, they analyze parts of the descriptions usually taking them into account with different weights. This motivates the use of algorithm models based on partial precedence.

Let β_ω be a partial similarity function on the pairs of descriptions

$$\beta_\omega : \prod_{t=1}^{s} \mathbf{M}_{i_t} \times \prod_{t=1}^{s} \mathbf{M}_{i_t} \longrightarrow V, \tag{3}$$

where $\omega = \{x_{i_1}, \dots, x_{i_s}\} \subseteq \mathbf{R}$, and V is as in (2).

Let β be a total similarity operator

$$\beta(\mathcal{O}_i, \mathcal{O}_j) = \frac{1}{|\Omega_A|} \sum_{\omega \in \Omega_A} \gamma_\omega \beta_\omega(\mathcal{O}_i, \mathcal{O}_j),$$

where γ_ω is a weighting parameter, denoting the informational importance of the subset of features ω. Ω_A is the family of subsets of features considered for the analysis of partial similarities.

Let $M = \{\mathcal{O}_1, \dots, \mathcal{O}_m\} \subseteq \mathbf{U}$. The only restriction imposed on M is the following: $\forall j = 1, \dots, r$ $(\max_{1 \le i \le m} \mu_j(\mathcal{O}_i) \ge 0.5)$, which means that for each class K_j, M has at least one object which is "closer" to belonging to K_j, than not to belonging to it.

The classical problem under these conditions would be formulated as follows.

> Given the description $D(\mathcal{O})$ of an object $\mathcal{O} \in (\mathbf{U} \backslash M)$, and the information regarding M, find the r-tuple of membership of $\mathcal{O}$. In order to do this, an algorithm $\mathcal{A}$ has to be devised such that $\mathcal{A}(D(\mathcal{O}_1), \alpha(\mathcal{O}_1), \dots, D(\mathcal{O}_m), \alpha(\mathcal{O}_m), D(\mathcal{O})) = (\alpha_1^A(\mathcal{O}), \dots, \alpha_r^A(\mathcal{O}))$, where $\alpha(\mathcal{O}_i) \in [0, 1]^r$ and $\alpha_j^A(\mathcal{O}) \in [0, 1]$. $\alpha_j^A(\mathcal{O})$ is the membership degree of the object $\mathcal{O}$ to class K_j estimated by the algorithm A. (4)

In this statement, $D(\mathcal{O}_1), \alpha(\mathcal{O}_1), \dots, D(\mathcal{O}_m), \alpha(\mathcal{O}_m)$ represents the information regarding M.

Starting from this general statement of the problem, different situations may arise.

- Classes $K_1, \dots, K_r$ may be crisp and disjoint, crisp and overlapped, forming a fuzzy r-partition [4], or a fuzzy β_0-partition, or a fuzzy β_0-cover as defined in [5].
- Features may be of different nature (numerical, logical, linguistical); missing values may appear for any feature.
- The comparison functions between descriptions (β), and between values of the same feature (φ_i), may be of a different nature (Boolean, k-valued, real, fuzzy).
- The output of the algorithm may be Boolean, k-valued, fuzzy, linguistic.

According to (4) different cases may be obtained. Among them, the supervised classification problem as in formulation (1), without fuzzifying sources is obtained. At the present state of the art, there are not algorithmic models to solve some of these situations.

A VOTING ALGORITHM MODEL IN FUZZY ENVIRONMENT

Algorithms for the "calculation of evaluations" or voting algorithms, were introduced by Zhuravlev [6]. The basic idea contained in this model, consists of the classification of objects based on determined partial precedents. This means that several parts of the object to be classified (of its description) are compared with the corresponding parts of the already classified objects forming M. Based on the behavior of the simiarity between these parts, some integrations are made by applying a set of rules.

An extension of the original ideas to fuzzy conditions is described below. The model is made through six stages which are as follows.

1. A support sets system is a not empty set of not empty subsets of the set of features. Every support set determines a subdescription of the objects, based on which, comparisons are made. We are going to allow the subsets in our set to be fuzzy, and we are also going to use a comparison criteria by features that may be infinite-valued.

   For example, the typical testor family of M may be taken as a support sets system [7]. A testor τ of M is a subset of features such that, taking into account only the variables in τ, no more similar objects appear in different classes than those already appearing when we considered all the features of **R**. A testor is typical if it is no longer a testor when we eliminate any of its variables. We may also consider the typical Goldman's fuzzy testors, which are a generalization of the above-mentioned ones (see, for example, [7]).
2. A similarity function allows us to quantify the similarity between two objects. Particularly, an object to be classified and those in the learning sample M can be compared, by using at each moment the corresponding subdescriptions, based on a support set.

   The general expression
$$\beta(\omega\mathcal{O},\omega\mathcal{O}_t) = |\omega|^{-1} \sum_{x_i\in\omega} \mu_\omega(x_i)\cdot\varphi_i(x_i(\mathcal{O}),x_i(\mathcal{O}_t))$$
   may be considered.

   In this expression, ω represents a support set, $\mathcal{O}$ is the object to be classified, $\mathcal{O}_t$ is the t^{th} object of M, $\omega\mathcal{O}_t$ expresses the subdescription of $\mathcal{O}_t$ considering only the features in ω, $|\omega|$ is the cardinality of ω, x_i represents the i^{th} feature, $\mu_\omega(x_i)$ is the membership degree of x_i to ω, and $\varphi_i(x_i(\mathcal{O}),x_i(\mathcal{O}_t))$ represents the (similarity) comparison between $\mathcal{O}$ and $\mathcal{O}_t$, regarding only the x_i feature. We consider that $\beta(\omega\mathcal{O},\omega\mathcal{O}_t)\in[0,1]$, for all ω, $\mathcal{O}$, $\mathcal{O}_t$.
3. The rule to evaluate the similarity objectwise for a fixed support set gives us the preliminary "votes". This rule does not necessarily agree with the result of the similarity function, because it is possible to introduce parameters related with each row and/or with the features present in the support set. For example, we can use an expression such as: $\Gamma_\omega(\mathcal{O},\mathcal{O}_t) = \gamma_\omega P(\mathcal{O}_t)\beta_\omega(\mathcal{O},\mathcal{O}_t)$, where γ_ω and $P(\mathcal{O}_t)$ are parameters associated to the set ω and to the object $\mathcal{O}_t$, respectively, for example, their informational weights.
4. A rule to evaluate the similarity classwise for a fixed support set, allows us to totalize the evaluations by object within a class, mantaining a fixed support set. It is a way of "scrutinizing" without making the support set change. For example, we may consider the average of the preceding evaluations objectwise of each class.
5. A rule to evaluate the similarity classwise for the entire support sets system also sumarizes the evaluations, but now based on the consideration of all the support sets. It constitutes the final step of the scrutiny, and it can be the final step of the algorithm if we were interested in knowing only the degree of similarity (membership) of the object $\mathcal{O}$ to each class. In general, we stop here in the fuzzy case.
6. The algorithm can eventually take an additional step: a decision rule. Based on the votes obtained, this rule produces a final assigment of the object $\mathcal{O}$ to one or more classes. This is the usual process when we are working with crisp sets, where the decision of locating or not locating the object in one or more classes is required. For instance, in those for which the membership degree satisfies a given condition; let us say that it has to exceed a certain threshold, or to be the maximum value, etc.

KORA-Ω SUPERVISED CLASSIFICATION ALGORITHM MODEL

The KORA-3 algorithm, introduced in [8], was devised for the solution of supervised classification problems in geosciences with two disjoint classes, for objects described in terms of Boolean variables without missing values. The underlying idea in the initial method is to classify new objects based on a learning sample according to the verification of some complex properties for each class. Those properties are formed from three feature values, which appear "sufficient times" in one of the classes and "sufficiently little" in the other one. The KORA-Ω algorithm introduced in [9] is a generalization of these ideas for the fuzzy case.

Now, let us consider a problem such as (4). Given a comparison criterion between values for each variable, and some partial similarity operators, such as (2) and (3), respectively. We will consider $V = [0, 1]$ in both cases.

DEFINITION. *Let Ω_A be a family of support sets, and $\omega = \{x_{i_1}, \dots, x_{i_p}\} \in \Omega_A$. A combination of values, $\mathbf{a} = (a_{i_1}, \dots, a_{i_p})$, and its respective features, forms a δ_i-fuzzy complex feature $(\mathbf{a}, \omega)$ for class K_i', with degree $\mu^i((\mathbf{a}, \omega))$, $i = 1, \dots, r$ iff*

1. *$\exists \mathcal{O}_j \in K_i' \omega K_i' \cap_\beta \{\mathbf{a}\} \supseteq \{(\mathcal{O}_j, \mathbf{a})\}$, where $A \cap_\beta B = \{(\mathcal{O}, \mathcal{O}') \mid \mathcal{O} \in A \wedge \mathcal{O}' \in B \wedge \beta(\mathcal{O}, \mathcal{O}') > 0\}$, being β a similarity function. In other words, this refers to the set of pairs of subdescriptions of objects, corresponding to the features $x_{i_1}, \dots, x_{i_p}$, which, according to β, have a similarity value strictly above zero. $\omega K_i'$ is the set of $\omega\mathcal{O}$ subdescriptions, with $\mathcal{O} \in K_i'$.*
2. *$\sum_{\mathcal{O}_j \in K_i'} \beta(\omega\mathcal{O}_j, \mathbf{a}) \mu_{K_i}(\mathcal{O}_j) \geq \delta_i$.*
3. *$\sum_{\mathcal{O}_j \in CK_i'} \beta(\omega\mathcal{O}_j, \mathbf{a})(1 - \mu_{K_i}(\mathcal{O}_j)) \leq \delta_i'$, being $CK_i' = \{\mathcal{O}_j \mid \alpha_i(\mathcal{O}_j) = 0\}$ with*

$$\mu^i((\mathbf{a}, \omega)) = \frac{\sum_{j=1}^m \beta(\omega\mathcal{O}_j, \mathbf{a}) \mu_{K_i}(\mathcal{O}_j)}{\sum_{j=1}^m \mu_{K_i}(\mathcal{O}_j)}.$$

$\delta_i \geq 0$,and $\delta_i' > 0$ are thresholds that denote to what degree $\mathbf{a}$ appears sufficiently often in objects of the K_i' class, and sufficiently seldom in the objects of the complement of that class, respectively. It may be demonstrated that, for ($\beta(\omega\mathcal{O}, \omega\mathcal{O}_j) = 1$ iff $\omega\mathcal{O} = \omega\mathcal{O}_j$) and $|\omega| = 3$, original KORA-3 algorithm is obtained.

Let $RC(K_i')$ be the set of all the δ_i-fuzzy complex features $(\mathbf{a}, \omega)$ for class K_i', with degree $\mu^i((\mathbf{a}, \omega))$, $i = 1, \dots, r$.

DEFINITION. *We will call η_i-remainder of K_i' to the set $r(K_i')$ of objects $\mathcal{O} \in K_i'$ such that $\sum_{(\mathbf{a}, \omega) \in RC(K_i')} \beta(\omega\mathcal{O}, \mathbf{a}) < \eta_i$, where η_i is a threshold value, $i = 1, \dots, r$.*

The η_i-remainder of K_i' contains those objects in K_i' that are not sufficiently represented in $RC(K_i')$.

DEFINITION. *A combination of values $\mathbf{a} = (a_{i_1}, \dots, a_{i_p})$ of the features $x_{i_1}, \dots, x_{i_p}$, respectively, forms a δ_{ij}-complementary fuzzy complex feature $(\mathbf{a}, \omega)$ for class K_i', with degree $\mu^i((\mathbf{a}, \omega))$, $i = 1, \dots, r$ iff*

1. *$\exists \mathcal{O} \in r(K_i') \omega r(K_i') \cap_\beta \{\mathbf{a}\} \supseteq \{(\mathcal{O}_j, \mathbf{a})\}$;*
2. *$\sum_{\mathcal{O} \in r(K_i')} \beta(\omega\mathcal{O}, \mathbf{a}) \geq \delta_{ij}$;*
3. *$\sum_{\mathcal{O} \in CK_i'} \beta(\omega\mathcal{O}, \mathbf{a}) \leq \delta_{ij}'$ with*

$$\mu^i((\mathbf{a}, \omega)) = \frac{\sum_{\mathcal{O} \in r(K_i')} \beta(\omega\mathcal{O}, \mathbf{a}) \mu_i(\mathcal{O})}{\sum_{\mathcal{O} \in r(K_i')} \mu_i(\mathcal{O})}.$$

$\delta_{ij} \geq 0$ and $\delta_{ij}' > 0$ are thresholds, that form decreasing monotonous successions of values, and which have the same meaning of the thresholds δ_i and δ_i', respectively; j is the iteration number.

KORA algorithm model defines complementary fuzzy complex features in order to cover the objects belonging to the remainders of the classes.

DEFINITION. *Let $(\mathbf{a},\omega)$ be a δ_i-fuzzy complex feature (complementary or not). We will call the magnitude written below, informational weight of the δ_i-fuzzy complex feature $(\mathbf{a},\omega)$:*

$$P((\mathbf{a},\omega)) = \sum_{x_i\in\omega} P(x_i) \sum_{\mathcal{O}_t\in OC(\mathbf{a},K'_j)} \gamma_j(\mathcal{O}_t),$$

where $P(x_i)$ and $\gamma_j(\mathcal{O}_t)$ are the informational weights of feature x_i and object $\mathcal{O}_t$ in class K'_j, respectively, calculated according to [7]; and $OC(\mathbf{a},K'_j)$ is the set of all the objects $\mathcal{O}\in K'_j$ such that $\omega\mathcal{O}=\mathbf{a}$.

The KORA Ω algorithm works in three stages.
Learning stage:

Step 1. determine δ_i, δ'_i, $i=1,2,\dots,r$, (by means of expert criteria);
Step 2. calculate all the δ_i-fuzzy complex features in each class.

Relearning stage:

Step 3. determine η_i and $r(K'_i)$, $i=1,2,\dots,r$;
Step 4. determine δ_{i1}, δ'_{i1}, $i=1,2,\dots,r$ (as in Step 1);
Step 5. calculate all the δ_i-fuzzy complex complementary features in each class.

The steps of this stage may be repeated as many times as required, changing the set $r(K'_i)$ in each iteration. In the worst case, it will be possible to make $|K'_i|$ iterations, $i=1,2,\dots,r$.
Classification stage:

Step 6. let $\omega\in\Omega_A$, and $(\mathbf{a}_p,\omega)$ be a δ_i-fuzzy complex feature (either complementary or not) for class K'_i, $p=1,\dots,s_i(\omega)$, where $s_i(\omega)$ is the number of δ_i-fuzzy complex feature (either complementary or not) for class K'_i with regard to $\omega\in\Omega_A$, $i=1,2,\dots,r$, and $\mathcal{O}$ an object to be classified; then:

$$\Gamma_i(\mathcal{O}) = \sum_{\omega\in\Omega_A}\sum_{p=1}^{s_i(\omega)} \left|\{\mathbf{a}_p\}\cap_\beta\{\omega\mathcal{O}\}\right| P\left((\mathbf{a}_p,\omega)\right);$$

Step 7. any rule to find the solution may be used; for example:
a) $\mathcal{O}\in K_i \Leftrightarrow \Gamma_i(\mathcal{O}) > \Gamma_j(\mathcal{O})$, $i\neq j$, $j=1,\dots,r$;
b) $\mathcal{O}\in K_i \Leftrightarrow \Gamma_i(\mathcal{O}) > \delta$, $i=1,\dots,r$, where δ is a threshold value.

CLASSIFICATION BASED ON SETS OF REPRESENTATIVES

The classification method based on the sets of representatives was introduced in [10] under the following conditions: objects are described in terms of a set of features that take values in metric spaces, classes are not necessarily disjoint, all comparison criteria are Boolean; the response of the algorithm will only take two extreme values, meaning for full belonging or not belonging. The basic underlying idea is that for each class some combinations of values of certain features may help to make a decision regarding, the classification of the object into the class. In practice, however, there are many problems, as those described in [1,11], which despite the fact that they resemble this basic idea do not meet all the conditions assumed by the aforementioned algorithm. This means that the analogy concepts used by the specialists should not be modeled through Boolean functions. The degree of reliability that the specialist has regarding the membership of the objects under study to any of the classes is not always absolute, but rather gradual, subjective; and they should not be modeled through crisp sets. The features in terms of which the objects are frequently described, are of different natures and, in the descriptions of the objects, the values of all the features are not always present. So the model should allow us to deal with partial descriptions of objects and with features of different natures. The above listed characteristics may occur in all of the possible combinations.

An extension of the Baskakova's model to real conditions of problems arising in geosciences, medicine, and other disciplines is exposed in [12].

Let us consider again a supervised classification problem such as the one described in (4). Given a comparison criterion between values of each variable, and partial similarity operators, such as (2) and (3), respectively. We consider $V = [0,1]$ in both cases.

DEFINITION. *Let Ω^j_A be a family of support sets for class K_j, and $\omega = \{x_{i_1}, \dots, x_{i_p}\} \in \Omega^j_A$. A combination $\mathbf{a} = (a_{i_1}, \dots, a_{i_p})$ of feature values is called a positive representative $(\mathbf{a}, \omega)$ for K_j with degree $\mu_j((\mathbf{a}, \omega))$, if the following conditions hold:*

1. $\exists\, \mathcal{O}_i \in K'_j \omega\mathcal{O}_i = \mathbf{a}$;
2. $-\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})\mu_j(\mathcal{O}_i) \geq \eta_j$;
3. $-\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})(1-\mu_j(\mathcal{O}_i)) < \delta_j$ *with*

$$\mu_j((\mathbf{a}, \omega)) = \frac{\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})\mu_j(\mathcal{O}_i)}{\sum_{i=1}^{m} \mu_j(\mathcal{O}_i)}.$$

Let ωM^1_j be the set of all positive representatives for K_j.

This means that a positive representative $(\mathbf{a}, \omega)$ is a combination of values, such that, it is sufficiently similar to corresponding subdescriptions "in K_j". Notice that each value of similarity is multiplied by the membership degree of the corresponding object to the class K_j. Condition 3 expresses an opposite constraint regarding the complement of K_j.

DEFINITION. *A combination $\mathbf{a}$ of feature values is called a negative representative for K_j with degree $\mu_j(\mathbf{a})$, if the following conditions are fulfilled:*

1. $\exists\, \mathcal{O}_i \in CK'_j \omega\mathcal{O}_i = \mathbf{a}$;
2. $-\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})(1-\mu_j(\mathcal{O}_i)) \geq \eta_j$;
3. $-\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})\mu_j(\mathcal{O}_i) < \delta_j$ *with*

$$\mu_j(\mathbf{a}) = \frac{\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})(1-\mu_j(\mathcal{O}_i))}{\sum_{i=1}^{m} (1-\mu_j(\mathcal{O}_i))}.$$

Let ωM^0_j be the set of all negative representatives for K_j.

Observe that if the classes are crisp sets, then $\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})\mu_j(\mathcal{O}_i) = \sum_{\mathcal{O}\in K'_j} \beta(\omega\mathcal{O}_i, \mathbf{a})$ and analogously, $\sum_{i=1}^{m} \beta(\omega\mathcal{O}_i, \mathbf{a})(1-\mu_j(\mathcal{O}_i)) = \sum_{\mathcal{O}\in CK'_j} \beta(\omega\mathcal{O}_i, \mathbf{a})$.

The classification algorithm based on the sets of representatives may be expressed as follows.

Step 1. Determine Ω^j_A, $\eta_j y CK'_j$, $j = 1, \dots, r$.

Step 2. Determine $P^j_1, \dots, P^j_n, \gamma^j_1, \dots, \gamma^j_m$, $j = 1, \dots, r$, the parameters associated with $x_1, \dots, x_n$ and $\mathcal{O}_1, \dots, \mathcal{O}_m$, respectively. These parameters usually represent the relevance of features and objects, respectively, in each class.

Step 3. Given an object $\mathcal{O}$ to be classified, calculate $\Gamma_j(\omega, \mathcal{O})$ for every $j = 1, \dots, r$ and for every $\omega = \{x_{i_1}, \dots, x_{i_s}\} \in \Omega^j_A$, as follows:

$$\Gamma_j(\omega\mathcal{O}) = \begin{cases} V^*_j, & \text{if } \omega\mathcal{O} \in \omega M^1_j, \\ V^{**}_j, & \text{if } \omega\mathcal{O} \in \omega M^0_j, \\ 0, & \text{in other cases,} \end{cases}$$

where

$$V^*_j = \frac{1}{s}\sum_{t=1}^{s} P^j_{i_t} \frac{1}{\rho} \sum_{i=1}^{m} \gamma^j_{v_i} \mu_j(\mathcal{O}_i)\beta(\omega\mathcal{O}, \omega\mathcal{O}_i), \qquad \rho = \sum_{i=1}^{m} \gamma^j_{v_i}\mu_j(\mathcal{O}_i),$$

$$V^{**}_j = -\frac{1}{s}\sum_{t=1}^{s} P^j_{i_t} \frac{1}{\tau} \sum_{i=1}^{m} \gamma^j_{v_i} (1-\mu_j(\mathcal{O}_i))\beta(\omega\mathcal{O}, \omega\mathcal{O}), \qquad \tau = \sum_{i=1}^{m} \gamma^j_{v_i} - \rho.$$

Step 4. The magnitudes $\Gamma_j^+(\omega\mathcal{O})$ and $\Gamma_j^-(\omega\mathcal{O})$ are calculated for $j = 1, \dots, r$, as follows:

$$\Gamma_j^+(\omega\mathcal{O}) = \sum_{\Gamma_j(\omega\mathcal{O})\geq 0} \Gamma_j(\omega\mathcal{O})\mu_j(\mathrm{a}),$$

$$\Gamma_j^-(\omega\mathcal{O}) = \sum_{\Gamma_j(\omega\mathcal{O})< 0} \Gamma_j(\omega\mathcal{O})\mu_j(\mathrm{a}).$$

Step 5. The total evaluation $\Gamma_j(\mathcal{O})$, is calculated for each class K_j, as follows:

$$\Gamma_j(\mathcal{O}) = \begin{cases} 0, & \text{if } \Gamma_j^+(\mathcal{O}) = 0 \text{ and } \Gamma_j^-(\mathcal{O}) = 0, \\ \dfrac{\Gamma_j^+(\mathcal{O}) + \Gamma_j^-(\mathcal{O})}{2\max\left\{|\Gamma_j^+(\mathcal{O})|, |\Gamma_j^-(\mathcal{O})|\right\}}, & \text{in other cases.} \end{cases}$$

It is not difficult to see that $\Gamma_j(\mathcal{O}) \in [-0.5, 0.5]$. Based on this evaluation, the membership function of class K_j will be estimated such that it will be a continuous increasing monotonous function, that will take values in $[0, \max\{\mu_j, (\mathcal{O}_i)\}]$ with $\mathcal{O}_i \in M$; $\mu_j(\mathcal{O}) = 0$ iff $\Gamma_j(\mathcal{O}) = -0.5$; $\mu_j(\mathcal{O}) = 0.5$ iff $\Gamma_j(\mathcal{O}) = 0$; $\mu_j(\mathcal{O}) > 0.5$ iff $\Gamma_j(\mathcal{O}) > 0$ and $\mu_j(\mathcal{O}) < 0.5$ iff $\Gamma_j(\mathcal{O}) < 0$. Therefore, the r-tuple of membership for $\mathcal{O}$, $\alpha(\mathcal{O}) = (\alpha_1(\mathcal{O}), \dots, \alpha_r(\mathcal{O}))$ shall be defined as

$$\mu_j(\mathcal{O}) = \begin{cases} 0.5 + \Gamma_j(\mathcal{O})\left(2\max\{\mu_j(\mathcal{O}_i)\} - 1\right), & \text{if } \Gamma_j(\mathcal{O}) \geq 0, \\ 0.5 + \Gamma_j(\mathcal{O}), & \text{if } \Gamma_j(\mathcal{O}) < 0. \end{cases}$$

In the above expression, if the total evaluation of an object is 0, then there is as much evidence indicating that it is in the class, as indicating the contrary. Therefore, its membership shall be 0.5. If the total evaluation is bigger than 0, then there is more evidence that it is in the class, so the membership value shall be higher than 0.5. On the other hand, if the total evaluation is smaller than 0, then the membership shall be smaller than 0.5. In case the total evaluation takes extreme values, membership also will.

APPLICATIONS

The models exposed here have been applied in some of their particular formulations to solve several problems in geosciences successfully: Gomez-Herrera *et al.* [1] provided a forecast gasopetroliferous map in determined the type of soil in Cuba, using voting algorithms. Simultaneously, Boolean variables as "presence of Methane anomaly", nominal features as "type of lithology", as well as numerical variables as "slopes, distances, gradients" were used in this problem. The data also presented missing values. Authors declared 80% of coincidences between theoretical and practical results on 48 drillholes. Keilis-Borok and Soloviev [13] discuss other applications in geosciences.

There are also reported some applications in medicine. For example, Ikramova [14] worked on prognosis of epidemic processes, complications of infectious hepatitis A and B, early diagnosis of bacterial dysentery, meningococcic infection, and respiratory diseases; and Dorodnitsina [15] applied voting algorithms to prognosis throat cancer complications.

It is important to point out that supervised classification problems often appear in eosciences and in medicine. In both areas, there appear quantitative features mixed with qualitative features, and it is also common in the presence of incomplete descriptions of several objects, that is, object descriptions with missing values. In this sense, the application of logical-combinatory models seems to be a more natural way for dealing with the characteristics of soft sciences problems.

CONCLUSIONS

The previous models conform a family of classification algorithms based on partial precedence, under the conditions posed in (4), that allow the solution of complex problems that frequently arise in sciences as medicine, the geosciences, and others. For example: differential diagnoses of illnesses, technical diagnoses of equipment, forecast of phenomena, etc. In such problems, quantitative and qualitative variables may be found mixed in object descriptions and even lacking information. The classes may be fuzzy subsets. These algorithms are a good alternative, from the methodologic point of view, for the solution of these problems. They will be improved by more extensive application to real problems.

REFERENCES

1. J. Gómez-Herrera *et al.*, Gasopetroliferous forecast in Cuban ophiolitic association using mathematical modeling, *Geofísica Internacional*, (in Spanish), **33** (3), 447–467, (1994).
2. V. Valev and Y.I. Zhuravlev, Integer-valued problems of transforming the training tables in k-valued code in pattern recognition problems, *Pattern Recognition* **24**, 283–288, (1991).
3. E. Cheremesina and J. Ruiz-Shulcloper, Methodological aspects about application of mathematical models for pattern recognition in soft sciences, *Revista Ciencias Matemáticas*, (in Spanish) **13** (2), 93–108, (1992).
4. E.H. Ruspini, A new approach to clustering, *Information and Control* **15**, 22–32, (1969).
5. J. Ruiz-Shulcloper and J.J. Montellano Ballesteros, A new model of fuzzy clustering algorithms, In *Proc. EUFIT'95*, **3** 1484–1498, (1995).
6. Yu.I. Zhuravlev and V.V. Nikiforov, Algorithms for recognition based on calculation of evaluations, *Kibernetika* (in Russian) **3**, Moscow, 1–11, (1971).
7. M. Lazo-Cortés and J. 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).
8. M.N. Bongard *et al.*, Solution of geological problems using recognition programs, *Sov. Geologia* (in Russian) **6**, (1963).
9. L.A. De la Vega-Doria, Extending KORA-3 algorithm to fuzzy environments, Thesis (in Spanish), CINVESTAV-IPN, México, (1994).
10. L.V. Baskakova and Y.I. Zhuravlev, An algorithm model for recognition with representative sets and support sets systems, *Zh. Vichislitielnoi Matematiki i Matematicheskoi Fiziki* (in Russian) **21** (5), 1264–1275, (1981).
11. J. Ruiz-Shulcloper *et al.*, PROGNOSIS and its applications in geology and geophysics, In *Proc. III Iberoamerican Congress on Artificial Intelligence IBERAMIA'92* (in Spanish), pp. 561–586, (1992).
12. J.A. Carrasco-Ochoa, Classifiers based on representative sets, Thesis (in Spanish), CINVESTAV-IPN, México, (1994).
13. V. Keilis-Borok and A. Soloviev, *Workshop on Non-Linear Dynamics and Earthquake Prediction*, International Centre for Theoretical Physics, Trieste, Italy. Nov.-Dec. 1991, (1991).
14. J.Z. Ikramova, *Algorithms of Recognition and Diagnosis* (in Russian), FAN, Tashkent, Uzbekistan, (1982).
15. V.V. Dorodnitsina, About the application of recognition and classification methods to the solution of medical diagnosis, In *Mathematical Methods in Pattern Recognition and Discrete Optimization* (in Russian), (Edited by Yu. I. Zhuravlev), pp. 33–42, Calculus Center of Russian Academy of Sciences, Moscow, (1990).