work, we will study the properties
nalization and conditioning.

lence intervals. *Journal of the Amer-*
6, 1983.
onstruct possibility measures from a
ation Processing and Management of
IPMU'96, 757–762, 1996.
ral. Probability intervals: A tool
urnal of Uncertainty, Fuzziness and
)4.
lembership functions i: Comparing
Journal of Approximate Reasoning,
f possibility-probability consistence.
7.
atistical data. *European Journal of*
nd necessity measures: towards a
uzzy Sets and Systems, 10(1):15–20,
ory: *An approach to computerized*
8.
ibility/probability transformations.
FSA'91) Congress, volume Mathe-
, 1991.
listic graphical models. In *Proceed-*
t Techniques and Soft Computing.
analysis of information-preserving
ossibilistic formulations of uncer-
ems, 20(2):143–176, 1992.
te confidence intervals for the bino-
atistical Association, 74(368):894–
iograms from interval data. *Int.*
97.
y transformations: A comparison.
:291–310, 1992.
nce. Princeton University Press,
sformations. *Fuzzy Sets and Sys-*
tions and their adquisition. *Fuzzy*
ry of possibility. *Fuzzy Sets and*

Fuzzy random variables-based modeling with GA-P algorithms

Luciano Sánchez
Departamento de Informática

Inés Couso
Dpto. Estadística e I.O. y D.M.

Universidad de Oviedo (España)
luciano@lsi.uniovi.es

Abstract: GA-P algorithms combine genetic programming and genetic algorithms to solve symbolic regression problems. In this work a fuzzy arithmetic-based GA-P procedure will be applied to the search of an analytic expression that relates input and output variables.

The algorithm has been tested in some practical problems, most of them related to electrical engineering. We will derive here an expression that relates climatic conditions with domestic electrical energy consumption in the north of Spain.

1 Introduction

Regression techniques (also named *modeling* techniques in machine learning literature) can be used to find an adequate expression for a function g so that given a variable Y that depends on the value of a variable X, $g(X)$ is a good approximation to Y. It is often assumed that the expression of g is known and that it depends on a set of parameters.

When the expression of g is not known previously and the structure of g and the values of the parameters on which it depends are to be determined, the technique is known as *symbolic regression*. There are many interval and fuzzy approaches to classical regression[9] but, as far as we know, symbolic regression methods find only *punctual*[6] and interval[14] estimates. These works will be extended here so that fuzzy arithmetic-based expressions can be obtained.

2 Interval predictions

When we need to model a physical process, we often want to know the value of a variable Y, using the information provided by a tuple of variables $X = (X_1, \ldots, X_d)$. We search for a function g such that the difference $Y - g(X)$ is small for every value of X; in other words a function g so that $g(X)$ is a good estimate of Y.

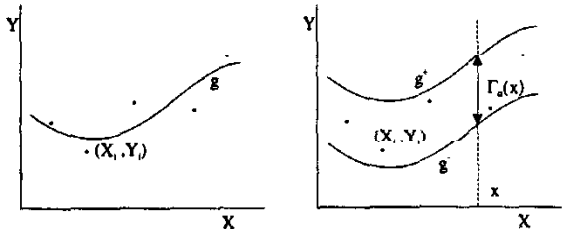

Figure 1: Punctual and interval regression. Interval models produce a range of values which contain the true outcome of the experiment with a high probability.

From a stochastic point of view we can assume that there exists a random experiment governed by a probability measure P with results in a set Ω such that $Y : \Omega \longrightarrow \mathbf{R}$ is a random variable and $X : \Omega \longrightarrow \mathbf{R}^d$ is a random vector. The function g that minimizes the mean square error in that case is $g(x) = E[Y \mid X = x]$ [10].

2.1 Multi-valued predictions

In some practical problems it is also interesting to obtain the margins in which we expect the variable Y is when the variable X (which can be multidimensional) is known. When we need to solve the punctual problem, we search for a function g such that $g(X)$ estimates $E[Y \mid X]$. Now we need an interval of values Γ_β that covers the value Y with probability higher than a confidence degree β and this interval must depend on the value of the variable X, so Γ_β is a function of X. Since X is also a function of ω, this mapping between the result ω of a random experiment and an interval $\Gamma_\beta(X(\omega))$ is a random set.

Formally, we will look for a multi-valued mapping $\Gamma_\beta : \mathrm{Im}(X) \longrightarrow I(\mathbf{R})$, where $I(\mathbf{R})$ is the set formed by all closed intervals in $\mathbf{R}$, such that the random set $\Gamma_\beta \circ X : \Omega \longrightarrow I(\mathbf{R})$ verifies

$$P\{\omega \in \Omega \mid Y(\omega) \in \Gamma_\beta \circ X(\omega)\} \geq \beta$$

for a given degree of confidence β (the symbol "$\circ$" means composition: $\Gamma_\beta \circ X(\omega) := \Gamma_\beta(X(\omega))$ and β is the probability that Y is in the interval $\Gamma_\beta(X)$.)

We can assess an interval prediction in some different ways. For example, we can say that given a value for β, the shorter Γ_β is, the better it is. Let us define two functions g^+ and g^- so that $g^-(X)$ is the minimum value of the confidence interval $\Gamma_\beta(X)$ and $g^-(X)$ is the maximum,

$$\Gamma_\beta \circ X = [g^- \circ X, g^+ \circ X]$$

and let us impose that g^+ and g^- are continuous (see Figure 1). Then, the margin of validity will be better when the mean difference between g^+ and g^-

$$E(g^+(X) - g^-(X))$$

is low. Since it must be true that

$$P\{\omega \in \Omega \mid g^-(X(\omega)) < Y(\omega) < g^+(X(\omega))\} \geq \beta$$

we can define the objective of the interval prediction as "find two functions g^+ and g^- such that the distance between $g^+(X)$ and $g^-(X)$ is minimum and Y is between $g^+(X)$ and $g^-(X)$ with a probability β".

In other words, given a region

$$R_{(g^+,g^-)} = \{(x,y) \in \mathbf{R}^{d+1} \mid g^-(x) < y < g^+(x)\}$$

we need to minimize

$$E(g^+(X) - g^-(X))$$

constrained to

$$P\{\omega \in \Omega \mid (X,Y)(\omega) \in R_{(g^+,g^-)}\} \geq \beta.$$

If we are solving an interval modeling problem we need to find two functions g^+ and g^- instead of the single function g that we needed to find in punctual modeling. Let us suppose now that we define g^+ and g^- by means of a function of X that depends on some interval parameters, using interval arithmetic [1]. This concept is similar to that introduced in [8] and many other works related to fuzzy regression [9]. Formally, let g^+ and g^- depend on a function $h_\theta : \mathbf{R}^m \longrightarrow \mathbf{R}$ so that $[g^-(x), g^+(x)] = \{t \in \mathbf{R} \mid t = h_\theta(x), \theta \in [\theta_1^-, \theta_1^+] \times \ldots \times [\theta_m^-, \theta_m^+]\}$ where the expression of h_θ is known except for the value of $2m$ parameters θ_k, $k = 1, \ldots 2m$ and h_θ is continuous with respect to θ and x (and then g^+ and g^- will also be continuous functions, as we had proposed). Given a function h, a random sample of size N obtained from the random vector (X,Y),

$$((X_1,Y_1),\ldots,(X_N,Y_N))$$

(where (X_i,Y_i) are independent and identically distributed) and a confidence degree $1 - \epsilon$ we can estimate $\theta_i^-(\epsilon)$ and $\theta_i^+(\epsilon)$ with the $2m$ values that minimize

$$\frac{1}{N}\sum_{i=1}^{N}(g^+(X_i) - g^-(X_i))$$

constrained by

$$1 - \epsilon \leq \frac{1}{N}\#\{i \in \{1 \ldots N\} \mid (X_i,Y_i) \in R_{(g^+,g^-)}\}$$

that is, the number of elements in the sample that belong to $R_{(g^+,g^-)}$.

For a given value of ϵ we can estimate the value of β by means of a second sample

$$((X'_1,Y'_1),\ldots,(X'_M,Y'_M)),$$

independent from the first one, by means of

$$\hat{\beta}_M = \frac{1}{M}\#\{i \in \{1 \ldots M\} \mid (X'_i,Y'_i) \in R_{(g^+,g^-)}\}.$$

The random variable $M \cdot \hat{\beta}_M$ follows a binomial distribution with parameters M and β and, by the strong law of the large numbers, it converges almost surely to the value β when $M \to \infty$. The procedure proposed should take three steps:

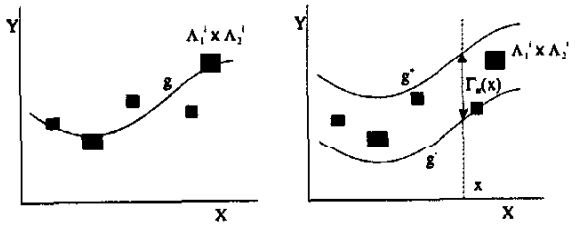

Figure 2: Punctual and interval estimation from imprecise data. Interval models are better suited for using interval data, since the error criterion for interval models can be applied directly to interval data.

1. Choose h and a (low) value for ϵ
2. Estimate the interval parameters $[\theta_i^-, \theta_i^+]$ from the first sample (training set)
3. Estimate the confidence degree β from the second sample (test set)

2.2 Imprecisely measured data

Let us suppose that the values of the random variable Y and the random vector X cannot be precisely observed but we only know that for a given output ω of the random experiment

$$(X,Y)(\omega) \in \Lambda(\omega)$$

where $\Lambda = \Lambda_1 \times \Lambda_2$, with $\Lambda_1 : \Omega \longrightarrow I(\mathbb{R}^d)$ and $\Lambda_2 : \Omega \longrightarrow I(\mathbb{R})$ are random sets, and $I(\mathbb{R}^d)$ is the set of all the rectangles in $\mathbb{R}^d$ (see Figure 2.) For example, imagine that we have a sensor that indicates "between 100 and 110" when $X(\omega_1) = 100$ and also when $X(\omega_2) = 105$; we model this behavior by means of a random set Λ_1 such that $\Lambda_1(\omega_1) = [100, 110]$ and $\Lambda_1(\omega_2) = [100, 110]$, and it is true that $X(\omega_1) \in \Lambda_1(\omega_1)$ and $X(\omega_2) \in \Lambda_1(\omega_2)$.

In these conditions, there is not an extension of the classical modeling that is universally accepted as the best one. But the previous model can deal with this imprecision. Let us define the two functions g^+, g^- so that $P\{\omega \in \Omega \mid g^-(X(\omega)) < Y(\omega) < g^+(X(\omega)) \quad \forall (X,Y) \in C(\Lambda)\} \geq \beta$ and let $C(\Lambda) = \{U \text{ random variable} \mid U(\omega) \in \Lambda(\omega) \quad \text{a.s. } (P)\}$ be the set of all random variables ***contained*** in Λ (see [3]). Then, the model we defined in the last section is valid for imprecise data. The set $C(\Lambda)$ includes all possible mappings (random variables) that can relate a result ω of the random experiment with values $X(\omega)$ and $Y(\omega)$ that are compatible with the imprecise observations Λ_1 and Λ_2.

We wish that the mean margin between g^+ and g^- is the narrowest possible one for a given β, but now it is posed an additional difficulty, because we do not know (X,Y) but a random set Λ which contains it. Taking a pessimistic criterion, we search for a region $R_{(g^+,g^-)} = \{(x,y) \in \mathbb{R}^{d+1} \mid g^-(x) < y < g^+(x)\}$ for which all points in the set $P_\Lambda(R_{(g^+,g^-)}) = \{P_{(X,Y)}(R_{(g^+,g^-)}) \mid (X,Y) \in C(\Lambda)\}$ $= \{t \in [0,1] \mid t = P\{\omega \in \Omega \mid g^- \circ X(\omega) < Y(\omega) < g^+ \circ X(\omega)\}, (X,Y) \in C(\Lambda)\}$

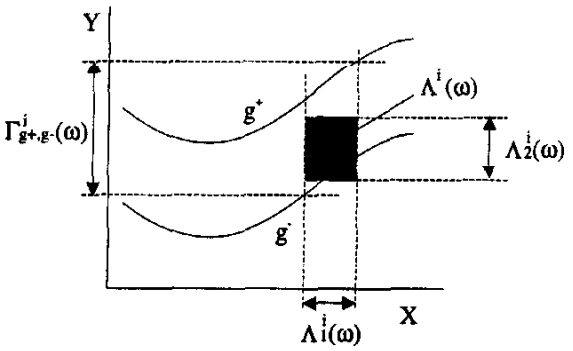

Figure 3: Calculations of the value of $\Gamma \circ \Lambda_X$. When the input value is an interval, the output is the projection of the intersection between the interval graph of the model and the cylindrical extension of the input.

are higher than the confidence β. For every pair of variables (X, Y) contained in Λ we obtain a value for the probability that (X, Y) is in $R_{(g^+,g^-)}$: the set $P_\Lambda(R_{(g^+,g^-)})$ is the set of all these values and it is bounded by the numbers β^- and β^+ [3] where

$$\beta^- = P\{\omega \in \Omega \mid \Lambda(\omega) \subset R_{(g^+,g^-)}\}$$

$$\beta^+ = P\{\omega \in \Omega \mid (\Lambda(\omega) \cap R_{(g^+,g^-)}) \neq \emptyset\}.$$

Making

$$\beta^- \leq \beta$$

so that

$$\inf_{(X,Y)\in C(\Lambda)} P\{\omega \in \Omega \mid g^- \circ X(\omega) < Y(\omega) < g^+ \circ X(\omega)\} \geq \beta$$

we obtain a model that fulfills that the probability that Y is in the interval prediction is higher that β in the worst case.

When data were precisely observed, we tried to minimize the expected length of the random interval $[g^-(X), g^+(X)]$ constrained by $P[\omega \in \Omega \mid g^-(X(\omega)) < Y(\omega) < g^+(X(\omega))] \geq \beta$. This time we want to find the minimum expected length of the random interval $\Gamma_{(g^+,g^-)}$ (see Figure 3)

$$\Gamma_{(g^+,g^-)}(\omega) = \{y \in \mathbf{R} \mid y \in [g^-(x), g^+(x)] \ \wedge \ x \in \Lambda_1(\omega)\}$$

$$= [\min_{x\in\Lambda_1(\omega)} g^-(x), \max_{x\in\Lambda_1(\omega)} g^+(x)]$$

(where the last assertion is true because g^- and g^+ are continuous functions and we know $\Gamma_{(g^+,g^-)}$ is strongly measurable by the same reason) restricted to

$$\beta \leq \min_{(X,Y)\in C(\Lambda)} P\{\omega \in \Omega \mid g^-(X(\omega)) < Y(\omega) < g^+(X(\omega))\}.$$

To solve the problem, we propose the following estimation: let

$$(\Lambda^1, \ldots, \Lambda^N) = (\Lambda_1^1 \times \Lambda_2^1, \ldots, \Lambda_1^N \times \Lambda_2^N)$$

be a size N random sample drawn from the random set $\Lambda = \Lambda_1 \times \Lambda_2$. For a given sample, we choose a value $\epsilon > 0$ as before and also a function h_θ known except for the values of m parameters, and we search for $2m$ constants θ_i^-, θ_i^+ so that $[g^-(x), g^+(x)] = \{t \in \mathbf{R} \mid t = h_\theta(x), \theta \in [\theta_1^-, \theta_1^+] \times \ldots \times [\theta_m^-, \theta_m^+]\}$ and the value $\widehat{L}_h = \frac{1}{N} \sum_{i=1}^N ||\Gamma^i_{(g^+,g^-)}||$ *is minimum*, where $\Gamma^i_{(g^+,g^-)} = \{y \in \mathbf{R} \mid y \in [g^-(x), g^+(x)] \wedge x \in \Lambda_1^i\}$ and restricting the search to the set of functions g^+, g^- that fulfill $1 - \epsilon \leq \frac{1}{N} \#\{i \in \{1 \ldots N\} \mid \Lambda^i \subset R_{(g^+,g^-)}\}$ that is, the number of intervals in the sample that are contained in $R_{(g^+,g^-)}$.

Once g^+ and g^- have been found, we cannot estimate β but a range of values for β if we have a second independent sample

$$(\Lambda'^1_1 \times \Lambda'^1_2, \ldots, \Lambda'^M_1 \times \Lambda'^M_2)$$

for which the value

$$\hat{\beta}_M^- = \frac{1}{M} \# \left\{ i \in \{1 \ldots M\} \mid \Lambda'^i \subset R_{(g^+,g^-)} \right\}$$

is an estimator of the *belief measure* [15] of the event "the imprecisely observed pair (X, Y) is in R", and

$$\hat{\beta}_M^+ = \frac{1}{M} \# \left\{ i \in \{1 \ldots M\} \mid \Lambda'^i \cap R_{(g^+,g^-)} \neq \emptyset \right\}.$$

is an estimation of the *plausibility* [15] of the same event. In other words,

$$\hat{\beta}_M^- \xrightarrow[M \to \infty]{a.s.} P[\omega \in \Omega \mid \Lambda(\omega) \subseteq R_{(g^+,g^-)}] =$$

$$\inf_{(X,Y) \in C(\Lambda)} P_{(X,Y)}(R_{(g^+,g^-)})$$

$$\hat{\beta}_M^+ \xrightarrow[M \to \infty]{a.s.} P[\omega \in \Omega \mid (\Lambda(\omega) \cap R_{(g^+,g^-)}) \neq \emptyset] =$$

$$\sup_{(X,Y) \in C(\Lambda)} P_{(X,Y)}(R_{(g^+,g^-)}).$$

Finally, note that the case analyzed in the previous section is a particular case of this one (where $\Lambda_1 = X$ and $\Lambda_2 = Y$).

3 Fuzzy predictions

In this section we will extend the technique so that it can be used to generate fuzzy models and process fuzzy data.

Let $(\widetilde{\Lambda}^1, \ldots, \widetilde{\Lambda}^N)$ be a tuple of realizations of a fuzzy random variable $\widetilde{\Lambda} : \Omega \longrightarrow \widetilde{\mathcal{P}}(\mathbf{R}^{d+1})$ that verifies that $(\widetilde{\Lambda})_\alpha : \Omega \longrightarrow I(\mathbf{R}^{d+1})$ is a random interval for all α, so we can say that $((\widetilde{\Lambda}^1)_\alpha, \ldots, (\widetilde{\Lambda}^N)_\alpha)$ is a random sample obtained from this random set. If we select a confidence degree β and apply the process

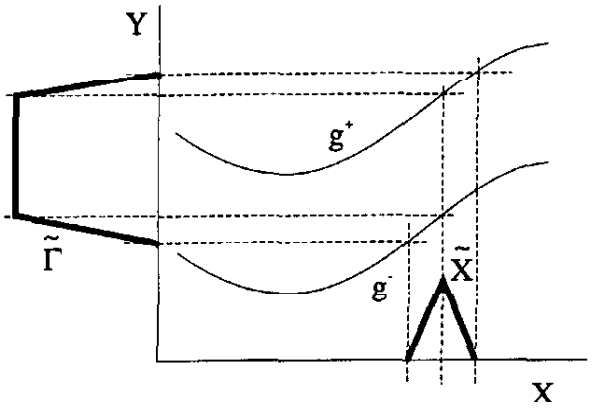

Figure 4: α-cuts of fuzzy models are interval models. The output of an interval model, when the input is a fuzzy set, is also the projection of the intersection between the interval graph of the model and the cylindrical extension of the input. This projection is a fuzzy set. The output of a fuzzy model is defined by the union of the outputs produced by every one of its α-cuts when the same input is applied to all of them.

shown before to an arbitrary α-cut α_0 of the sample $((\tilde{\Lambda}^1)_{\alpha_0}, \ldots, (\tilde{\Lambda}^N)_{\alpha_0})$ we will obtain two functions g_β^+ and g_β^- that fulfill $\min\{P\{\omega \in \Omega \mid g_\beta^- \circ X(\omega) < Y(\omega) < g_\beta^+ \circ X(\omega)\} \mid (X,Y) \in C((\tilde{\Lambda})_{\alpha_0})\} \geq \beta$.

Let us suppose that the fuzzy observation $\tilde{X}$ is a Cartesian product of fuzzy numbers

$$\tilde{X} = \tilde{X}_1 \times \ldots \times \tilde{X}_n$$

where the membership function of $\tilde{X}$ is $\tilde{X}(x_1, \ldots, x_n) = \tilde{X}_1(x_1) \wedge \ldots \wedge \tilde{X}_n(x_n)$ so the α-cuts of the fuzzy measurements are also intervals in $\mathbb{R}^d$, $\tilde{X}_\alpha \in I(\mathbb{R}^d)$ for all $\alpha \in [0,1]$. The multi-valued prediction of the interval model for an α-cut of $\tilde{X}$ is

$$\Gamma^\alpha_{(g^+,g^-)}(\beta) = [\min_{x \in \tilde{X}_\alpha} g_\beta^-(x), \max_{x \in \tilde{X}_\alpha} g_\beta^+(x)]$$

and it can be seen that

$$\alpha_1 < \alpha_2 \Rightarrow \Gamma^{\alpha_1}_{(g^+,g^-)}(\beta) \subseteq \Gamma^{\alpha_2}_{(g^+,g^-)}(\beta)$$

and, since g^+ and g^- are continuous, the family of intervals $\{\Gamma^\alpha_{(g^+,g^-)}(\beta)\}_{\alpha \in [0,1]}$ define a fuzzy set $\tilde{\Gamma}_{(g^+,g^-)}(\beta)$ whose membership function is

$$\tilde{\Gamma}_{(g^+,g^-)}(\beta, x) = \sup\{\alpha \mid x \in \Gamma^\alpha_{(g^+,g^-)}(\beta)\}.$$

A graphic example of this partial result can be seen in figure 4.

If we repeat the process for a certain rank of values of β we obtain a family of regions $\{R_{(g^+,g^-)}(\beta)\}_{\beta \in [0,1]}$. Under certain conditions this family can be used to construct a fuzzy set $\tilde{R}(x,y) = \sup\{\alpha \mid (x,y) \in R_{(g^+,g^-)}(\beta(\alpha))\}$.

In particular, for continuous random variables it makes sense to define every set $R_{(g^+,g^-)}$ by means of cuts of the density function of the random vector (X,Y) with horizontal planes. If the density function of (X,Y) is continuous and the density function of Y conditioned to X is unimodal, these sets have the minimum area, are nested and they form a fuzzy set (we do not include the demonstration here, because of limitations of space).

The mentioned process of calculating the output of a fuzzy model can also be written as follows:

$$\tilde{Y}(y) = \bigvee_{x \in \mathbf{R}^d} (\tilde{R}(x,y) \wedge \tilde{X}(x))$$

where $\wedge$ and $\vee$ mean "minimum" and "maximum" and $\tilde{R} : \mathbf{R}^{d+1} \longrightarrow [0,1]$ is the *fuzzy graph* of the model. Notice the similarities between this expression and the inference in an approximative Mamdani rule-based model.

4 Numerical optimization method. GA-P Algorithms

GA-P algorithms are an evolutionary computation method, hybrid between genetic algorithms and genetic programming, optimized to perform symbolic regressions. A complete description of the GA-P method can be found in [5]. Briefly, we will remark that each element of the population comprises a chain of parameters and a tree that describes a function, which depends on these parameters. The two operations by means of which new members of the population are generated are crossover and mutation. In a GA-P algorithm both operations are independently performed over the tree and the chain of parameters; i.e., we can cross or mutate only the chain, only the tree, both or none of them. The object of the method is to concurrently evolve the tree and the chain of parameters.

4.1 Modifications to GA-P

We will implement the fuzzy model proposed in the last section by means of a function h_θ that depends on m fuzzy parameters $\tilde{\theta}_i$. The terminal nodes of the tree that codifies h_θ must be fuzzy sets and fuzzy arithmetic operators are used to evaluate the tree; apart from this, modifications to Interval GA-P[14] are straightforward.

As in interval GA-P, the fitness does not depend on the evaluations of the tree part in the set of examples but on the separation between g_β^+ and g_β^- for all β. We measure that separation with the non-specificity of the fuzzy graph $\tilde{R}$. Since every α-cut of $\tilde{R}$ is a confidence interval and we must maximize the confidence degree of all of them simultaneously, the following fitness function is used:

$$\begin{cases} f_1 = K \cdot \frac{1}{N}\sum_{i=1}^{N}[D(\tilde{Y}^i) - D(\tilde{Y}(\tilde{X}_i))] & \text{if } P_c < 1-\epsilon_1 \\ f_2 = \frac{1}{N}\sum_{i=1}^{N} U(\tilde{Y}(\tilde{X}_i)) & \text{if } P_c > 1-\epsilon_2 \\ f_3 = \frac{f_1\cdot(\epsilon_2+P_c-1)+f_2\cdot(1-\epsilon_1-P_c)}{\epsilon_2-\epsilon_1} & \text{else} \end{cases}$$

where $\tilde{Y}(\tilde{X}_i)$ is the fuzzy output of the model for an input $\tilde{X}_i$,

$$\tilde{Y}(\tilde{X}_i)(y) = \bigvee_{x\in\mathbf{R}^d} (\tilde{R}(x,y) \wedge \tilde{X}_i(x))$$

P_c is a measure of the degree of the covering of the sample by the fuzzy model,

$$P_c = \frac{1}{N}\#\{i \in \{1\ldots N\} \mid \text{supp}(\tilde{Y}_i) \subseteq \text{supp}(Y(\tilde{X}_i))\})$$

K is a real value high enough so that $K \cdot \frac{1}{N}\sum_{i=1}^{N}[D(\tilde{Y}^i) - D(\tilde{Y}(\tilde{X}_i))]$ is always higher than $\frac{1}{N}\sum_{i=1}^{N} U(\tilde{Y}(X_i))$, U is a measure of non-specificity and D is a defuzzification operator.

The explanation of this function follows: when in the initial stages of the evolution, every imprecise measurement is replaced by one point (the measures are "defuzzified") and the least squares solution is tracked. As soon as a fraction $1-\epsilon_1$ of the sample is covered, we begin to promote those individuals that get a narrower band of prediction values. We define the fitness function so that it guarantees that a model that covers a fraction higher or equal than $1-\epsilon_2$ of the examples is always better than a model that covers less than $1-\epsilon_1$, so population evolves gradually towards models with adequate covering that will be selected on the basis of their non-specificity.

We penalize the solutions that do not cover a fraction $1-\epsilon_2$ by multiplying their fitness by a value K that is determined empirically. We think that it is not necessary to resort to multi-criteria optimization (see [4]) because the value of K is not difficult to obtain. When the fraction of errors $\epsilon_1 < \epsilon_2$ is reached, the fitness is the non-specificity. In intermediate situations (covering between $1-\epsilon_2$ and $1-\epsilon_1$) the fitness ranges between the values of the non-specificity and the scaled square error.

5 Practical application

Fuzzy GA-P was recently applied in practice for determining the dependence between the demand of electrical energy and some climate conditions in the north of Spain. The company "Hidroeléctrica del Cantábrico, S.A." asked us to build one model that relates mean temperature in Asturias and the number of hours with sun light with the total consumption of electrical energy by the domestic customers. This model has a double purpose; first, they will use it for making predictions of the domestic demand and then planning the optimal use of the resources of the company and second, it will allow to gain insight into the behavior of the electrical market, which is going to be liberalized in the next months.

Method	Train (70%)	Test (30%)	Complexity
Linear	586733	1021640	4 parms., 3 variables
2th order poly.	566460	964543	10 parms., 3 variables
GA-P	587317	928633	3 parms., 1 variable
Fuzzy GA-P	500436	924122	2 fuzzy parms., 1 variable (95% of points)
MLP 3-10-1	531952	807688	55 parms.
Fuzzy Rule Based	586733	1021640	1 TSK rule

Table 1: Cross-comparison of results. The dataset is very noisy and there is not a definite gain when using non linear models. GA-P and Fuzzy GA-P models depend only on one of the three variables, the minimum temperature. The Fuzzy Rule-based model degenerated to a a linear model. Fuzzy GA-P error was estimated over a subset of 95% of the points defined by one of its α-cuts.

The set of data consists of 1096 points, comprising 12 measurements each. Three attributes are symbolic (day of week, bank holiday or not, name of month) and all the remaining ones (minimum and maximum temperatures along one day in three points of the area, and the number of minutes of sun in these three points) are numerical. We will not include the symbolic attributes as inputs to our models but we will use them to segment the dataset.

We will only show here the simplest of these models, that which globally relates temperatures and the amount of sun light with the global demand. The input variables are reduced to three: mean lowest temperature, mean highest temperature and mean number of light hours. First, a linear model was calculated. After that, a non linear least squares polynomial model was adjusted with Levenberg-Marquardt method. Third order polynomials were not studied because of the excessive number of parameters. Later, GA-P and fuzzy GA-P were launched. A fuzzy rule-based model and a multi-layer perceptron were also tried, for comparison purposes.

When running the genetic procedures, steady state approach, with tournament selection and elitism were used. The probability of crossover is 0.9, both in GP and GA parts. We did not perform mutation in the tree part and we applied this operator with probability 0.01 in GA part, which is encoded in floating point. We used local optimization (Nelder and Mead's simplex) and over selection (1000 individuals). The population had 10 niches with 100 individuals each. In Fuzzy GA-P we used a triangular parameterization of all fuzzy sets, and defined the defuzzification operator to be the point of maximum membership.

This dataset is very noisy, and there are a high number of outliers. Fuzzy GA-P probed to be rather insensitive to the effect of outliers. It is remarkable that both genetic procedures discarded two of the three variables and coincide in making the demand to depend only on the minimum temperature in the day. This dependence is non linear, and it was nor detected when the correlations between the output and every input were calculated.

6 Conc

GA-P method
is easy and ver
of the expressi
operations. T
methods able
classical regre

In this wo
arithmetic-bas
and parameter
mation is codi

Referenc

[1] Bojadziev
1995.

[2] Cordon, (
lutionary
1999.

[3] Couso, I.
de Invest
1997.

[4] Fonseca,
Multiobje

[5] Howard,
Programi

[6] Koza, J.

[7] Iba, H.,S.
programr
406. vol

[8] Ishibuchi
with inte
Fuzzy Se

[9] Kacprzyk

[10] Ljung, L

[11] Renders,
climbing
IEEE Cc

6 Conclusions

GA-P methods can discover an empirical law from a set of samples. The method is easy and very flexible, because it allows us to select the maximum complexity of the expression, the maximum number of parameters and an arbitrary set of operations. Therefore, GA-P's are very convenient when compared to other methods able to make this kind of study: trial and error, neural networks or classical regression.

In this work we have adapted the GA-P algorithm to produce a fuzzy arithmetic-based model. Our method produces fuzzy estimations for outputs and parameters, and it would allow to use symbolic information if that information is codified by mean of fuzzy sets.

References

[1] Bojadziev, G. *Fuzzy Sets, Fuzzy Logic, Applications.* World Scientific. 1995.

[2] Cordon. O. "Solving Electrical Distribution Problems Using Hybrid Evolutionary Data Analysis Techniques". *Applied Intelligence*, (**10**), pp 5-24. 1999.

[3] Couso, I. *La Envolvente Probabilística. Definición y Propiedades.* Trabajo de Investigación. Universidad de Oviedo. Departamento de Estadística. 1997.

[4] Fonseca, C., Fleming, P.J. "An Overview of Evolutionary Algorithms in Multiobjective Optimization". Evolutionary Computation 3, 1-16. 1995.

[5] Howard, L.; D'Angelo, D. "The GA-P: A Genetic Algorithm and Genetic Programming Hybrid" IEEE Expert. June 1995. 11-15. 1995.

[6] Koza, J. R. *Genetic Programming.* MIT Press. 1992

[7] Iba, H.,Sato, T. De Garis, H. "System Identification approach to genetic programming". Proc First IEEE Conf on Evolutionary Computation. 401-406. vol 1. 1994.

[8] Ishibuchi, H., Tanaka, H., Okada, H. "An architecture of neural networks with interval weights and its application to fuzzy regression analysis". Fuzzy Sets and Systems 57. 27-39. 1993.

[9] Kacprzyk, J. *Fuzzy Regression Analysis.* Omnitech Press, Warsaw. 1992.

[10] Ljung, L. *System Identification: Theory for the User* Prentice Hall. 1987.

[11] Renders, J.M.; Bersini, H. "Hybridizing genetic algorithms with hill-climbing methods for global optimization: two possible ways" Proc. first IEEE Conf. Evolutionary Computation. 312-317, vol. 1. 1994.

[12] Renders, J.M., Flasse, S. P. "Hybrid Methods Using Genetic Algorithms for Global Optimization". IEEE Transactions on SMC. Part B: Cybernetics. Vol 26, NO. 2, April 1996.

[13] Sánchez, L. "A Random Sets Based Method for Identifying Fuzzy Models". Fuzzy Sets and Systems 98 (**3**). pp 343-354. 1998.

[14] Sánchez, L. "Interval-valued GA-P Algorithms". To appear in IEEE Transactions on Evolutionary Computation.

[15] Shafer *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey. 1986.