A Multiobjective Genetic Fuzzy System with Imprecise Probability Fitness for Vague Data

Luciano Sánchez, Inés Couso and Jorge Casillas

Abstract— When questionnaires are designed, each factor under study can be assigned a set of different items. The answers to these questions must be merged in order to obtain the level of that input. Therefore, it is typical for data acquired from questionnaires that each of the inputs and outputs are not numbers, but sets of values.

In this paper, we represent the information contained in such a set of values by means of a fuzzy number. A fuzzy statistics-based interpretation of the semantic of a fuzzy set will be used for this purpose, as we will consider that this fuzzy number is a nested family of confidence intervals for the value of the variable. The accuracy of the model will be expressed by means of an interval-valued function, derived from a recent definition of the variance of a fuzzy random variable.

A multicriteria genetic learning algorithm, able to optimize this interval-valued function, is proposed. As an example of the application of this algorithm, a practical problem of modeling in marketing is solved.

I. INTRODUCTION

When acquiring data from interviews or questionnaires, it is a common practice to evaluate the level of a factor by examining the answers to a set of different questions. These questions may show different aspects of the problem, or be redundant, to ensure the coherence of the data. Therefore, it is typical for data acquired from questionnaires that each of the inputs and outputs are not single numbers, but sets of values. The classical conversion of such a set of items into a compound value [1] consists in replacing each one of them by a suitable, numerical characteristic value, say its mean or median. This solution might not be the best one, because it discards the information about the dispersion of the items.

In this paper, we will transform each set of inputs into a fuzzy interval, that contains information about the average value and the dispersion of the items, and then learn the model from the produced fuzzy data. We will use a novel interpretation of the semantics of a fuzzy set [2], that regards fuzzy sets as families of confidence intervals, and let the learning be grounded in the minimization of an imprecisely known function, as proposed in [3]. We also propose to estimate the distance between the imprecisely known output data and the response of the model by mean of a new definition of squared error, derived from that of the variance of a fuzzy random variable. This paper also details the modifications that must be done to the NSGA-2 algorithm [4] in order

L. Sánchez is with the Computer Science Department of the Oviedo University, Spain. Email: luciano@uniovi.es. Inés Couso is with Statistics Department, Oviedo University, Spain. Email: couso@uniovi.es. Jorge Casillas is with the Computer Science Department, Granada University, Spain. Email: casillas@decsai.ugr.es.

to use our criterion. By last, this algorithm is applied to a marketing problem, that of modeling the consumer behavior from data obtained by questionnaires, and their crisp and fuzzy implementations are compared.

This work is organized as follows. Section II introduces the fuzzy representation issues, and Section III describes the imprecise probabilities-based objective criteria. Section IV briefly describes the practical problem based on consumer behavior models. Section V shows some obtained experimental results. Finally, Section VI concludes.

II. A FUZZY INTERPRETATION OF THE SEMANTIC OF A LIST OF VALUES

A. Semantics of a Fuzzy Set

Under the imprecise probabilities framework, it makes sense to understand a fuzzy set as a set of tolerances, each one of them is assigned a confidence degree, being the lower degree the narrower tolerance [5]. In particular, the α-cuts of the fuzzy set can be regarded as confidence intervals with degree $1 - \alpha$ [2].

This representation allows us to codify the information contained in a set of numbers by means of a fuzzy set. This will be made clear with the example that follows. Let us suppose that a variable X has associated the items valued

$$X = \{2, 1, 3, 3, 2, 2, 4\}. \tag{1}$$

The most immediate calculation of a summary value is the sample mean, which is 2.429. While this is a good compromise value, we are discarding information that might be relevant: there are some items as low as 1, and others as high as 4. To gain additional insight about the importance of the dispersion of the values, we will assume that the set of items X is a sample of a larger population, whose mean is unknown. Given the sample X, we can calculate confidence intervals for the value of this mean, at different degrees. If we want to simplify the calculations, we can assume that the sample was drawn from a normal population. Then, the α-cuts of the fuzzy set $\widetilde{X}$ that represent the value of the variable are the confidence intervals

$$\widetilde{X}_\alpha = 2.429 \pm 0.9759 \cdot \text{qt}_6 \left(1 - \frac{\alpha}{2}\right). \tag{2}$$

where qt_6 is the quantile function for the t distribution. A graphical representation of the membership function of $\widetilde{X}$ is shown in Figure 1. Observe that we can approximate it by a triangular membership function without incurring large errors. The same procedure must be applied to all lists of input and
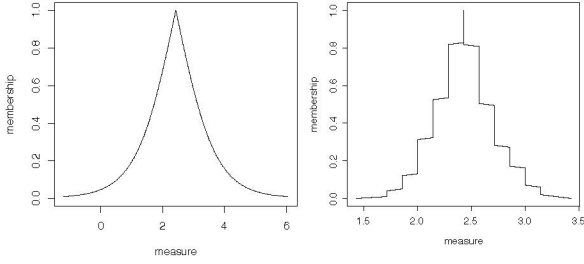
Fig. 1. Membership function of that set $\tilde{X}$ that represents the sample X in Section II-A. The left one was obtained under normality assumptions, and the right one is a basic bootstrap estimation.

output values, to obtain a fuzzy dataset, from which we want to learn a model. This will be discussed in the next section.

It is remarked that other, different techniques for estimating the needed confidence degrees would also be possible. If the sample mean is selected as the summary measure, the normality assumption will hold for many practical problems. Otherwise, when the distribution of the data is asymmetric, building the membership function from the quantiles of the bootstrap distribution of the mean or the median is a sensible choice. In the right part of Figure 1 we have plotted a bootstrap-based estimation of the membership function associated to the same data.

III. DESIGN OF LINGUISTIC MODELS FROM FUZZY DATA

A. Obtaining an output from a fuzzy input

Let our model f be a function that maps an input vector $\mathbf{x}$ to an output value $f(\mathbf{x})$. If all our knowledge about $\mathbf{x}$ is given by a fuzzy set $\tilde{X}$, built as described in the preceding section, then $y = f(\mathbf{x})$ is contained in the fuzzy set $f(\tilde{X})$, defined by the family of α-cuts

$$f(\tilde{X})_\alpha = \{f(\mathbf{x}) \mid \mathbf{x} \in \tilde{X}_\alpha\}. \tag{3}$$

Depending on the properties of f, the computation of $f(\tilde{X})$ can be NP-hard [6]. There exist some numerical procedures to solve this problem, many of them derived from the so-called vertex method [7]. We will use one of them, the fuzzy profile method [8] and approximate all the fuzzy sets by means of piecewise linear continuous membership functions, assuming that the supports of the input data are small enough so that our model f is locally monotonic with respect to each argument.

B. Measuring the quality of a model: different definitions of the Mean Squared Error

To select the best model, we want to measure the differences between its output and the desired values. Since we cannot know the precise desired output data, but fuzzy sets that describe them, we can not compute a number that measures the error of a candidate model over our train data, but we can provide a fuzzy interval for it. Let $\tilde{D} = \tilde{Y} - f(\tilde{X})$ be the residual of the model. We are interested in computing $E(\tilde{D}^2)$, and the definition of this value is similar to that of the variance of a frv. In the following sections, we define the mean squared error of three different forms, according to the three types of variance of a fuzzy random variable described in [9].

1) Classical model: Let us consider a probability space, $(\Omega, \mathcal{A}, P)$, and a metric, d, defined over the class of the fuzzy subsets of $I\!R$ (or over a subclass) and let us suppose that $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(I\!R)$ is a function $\mathcal{A} - \beta(d)$-measurable (here, $\beta(d)$ represents the Borel σ-algebra induced by d.) The *classical variance* of $\tilde{X}$ is the quantity

$$\mathrm{Var}_{\mathrm{Cl}}(\tilde{X}) = \int_\Omega d(X, E(\tilde{X}))^2 \, dP$$

and we will call *classical mean squared error (CMSE)* of a model with residual $\tilde{D}$ to the quantity

$$\mathrm{CMSE}(\tilde{D}) = \int_\Omega d(\tilde{D}, 0)^2 \, dP$$

The different definitions of variance in the literature that fit this formulation differ in the used metric and in the definition of the expectation of a fuzzy random variable [10][11], and the same distances would produce the corresponding definitions of CMSE. This definition, beside being the most convenient from a numerical point of view (the error of a fuzzy model would be crisp number, that could be easily optimized) is not compatible with our semantic interpretation, as we point out in the example that follows:

Example 3.1: Let us suppose that we have a sample of size 2 of the residual of the model, and that this residual is the frv $\tilde{D}$, whose images are the triangular fuzzy sets $\tilde{D}(\mathbf{x}_1) = (-K, 0, K)$ and $\tilde{D}(\mathbf{x}_2) = (-K, 0, K)$ (see Figure 2.) This frv, if regarded as a classical measurable function, have null CMSE, since it is a constant function. But, it is not coherent to say that we do not know the precise output of a model but a set that contains it and then, using only the information contained in this set, to assure that the unknown output has null variance. At most, we could know some bounds of this variance, because the margins of a function being constant do not mean that the function is a constant.

2) Imprecise 2nd order model: In [12], Kruse defines the variance of a multi-valued mapping, $\Gamma : \Omega \to \mathcal{P}(I\!R)$, as the set:

$$\mathrm{Var}_{\mathrm{Kr}}(\Gamma) = \{\mathrm{Var}(X) \mid X \in S(\Gamma)\},$$

where $S(\Gamma)$ represents the set of all measurable selections of the multi-valued mapping. The preceding definition can be easily extended to the case of fuzzy random variables as follows: Let us call *Kruse's variance* of the fuzzy random variable $\tilde{X} : \Omega \to \tilde{\mathcal{P}}(I\!R)$, to the only fuzzy set determined by the nested family of sets:

$$[\mathrm{Var}_{\mathrm{Kr}}(\tilde{X})]_\alpha := \mathrm{Var}_{\mathrm{Kr}}(\tilde{X}_\alpha), \forall \alpha,$$

where $\tilde{X}_\alpha$ is the multi-valued mapping α-cut of $\tilde{X}$. (The variance of the example 3.1 is plotted in the central part of the figure 2.)

Therefore, we define the *second order mean squared error (SMSE)* as the fuzzy set

$$[\mathrm{SMSE}(\tilde{D})]_\alpha := \mathrm{Sq}_{\mathrm{Kr}}(\tilde{D}_\alpha), \forall \alpha,$$

where

$$\mathrm{Sq}_{\mathrm{Kr}}(\Gamma) = \{\mathrm{E}(D^2) \mid D \in S(\tilde{D})\}.$$

We can easily check that the membership function of this fuzzy set is given by the expression

$$\text{SMSE}(\tilde{D})(x) = \sup\{\text{acc}(D) \mid \text{E}(D^2) = d\}, \ \forall \, d \in I\!R.$$

The membership degree of a value x to the fuzzy set $\text{SMSE}(\tilde{D})$ represents the possibility degree of the original random variable is one of those whose squared error is equal to x. This definition is conceptually interesting, but hard to include in practical problems without recurring to fuzzy rankings, or other heuristically defined order relations between fuzzy sets, which are not less arbitrary than the *classical* definition of the variance of a frv.

3) First order imprecise model: In our opinion, the most interesting definition, from a Genetic Learning point of view, is the one that follows. It produces an interval instead of a point or a fuzzy set, while keeps the coherence with the representation of the data.

We will consider, on the one hand, the probability measure P (defined over $\mathcal{A}$), that models the first sub-experiment, and, on the other, a family of conditional possibility measures, $\{\Pi(\cdot \mid \omega)\}_{\omega \in \Omega}$, defined as follows:

$$\Pi(A \mid \omega) = \Pi_{\tilde{X}(\omega)}(A) = \sup_{x \in A} \tilde{X}(\omega)(x), \ \forall \, A \in \beta_{I\!R}, \ \forall \, \omega.$$

In the preceding formula, $\Pi_{\tilde{X}(\omega)}$ represents the possibility measure determined by the possibility distribution $\tilde{X}(\omega)$: $I\!R \rightarrow [0,1]$. So, the value $\Pi(A|\omega)$ is an upper bound for the probability of the final outcome is in A, verifying the hypothesis that the outcome of the initial experiment is ω. This family of possibility measures represents our (imprecise) knowledge carried by $\tilde{X}$ about the relation that exists between the outcome of the first sub-experiment and the set of all the possible outcomes of the second one. Within this context, all we know about the probability distribution that models the second experiment is that it is in the set:

$$\{Q_2 \mid Q_2 \text{ marginal of } P \text{ and } Q(\cdot|\cdot), \ Q(\cdot|\cdot) \in \mathcal{C}\}$$

where $\mathcal{C} = \{Q(\cdot|\cdot) \text{ transition prob.} \mid Q(A|\omega) \leq \Pi(A|\omega)$

$$\forall \, A \in \beta_{I\!R}, \omega \in \Omega\}. \quad (4)$$

(The first order variance of the example 3.1 is plotted in the right part of the Figure 2.) Therefore, in the proposed imprecise probabilities model proposed, all we know about the squared error of the output of the second sub-experiment, or *first order mean squared error (FMSE)* is that it is in the interval:

$$\text{FMSE}(\tilde{D}) = \left\{ \begin{array}{l} \int_{I\!R} x^2 dQ_2 \mid Q_2 \text{ marginal of } P \\ \text{and } Q(\cdot|\cdot), \ Q(\cdot|\cdot) \in \mathcal{C} \end{array} \right\} \quad (5)$$

or, alternatively [9]:

$$\text{FMSE}(\tilde{D}) = \left\{ \int_{I\!R} d^2 dQ \mid Q(A) \leq \overline{P}_{\tilde{D}}(A), \ \forall \, A \in \beta_{I\!R} \right\} \quad (6)$$

where $\overline{P}_{\tilde{D}}$ is the sub-additive set-valued function given by

$$\overline{P}_{\tilde{D}}(A) = \int_0^1 P^*_{\tilde{D}_\alpha}(A) \, d\alpha, \forall \, A \in \beta_{I\!R}$$
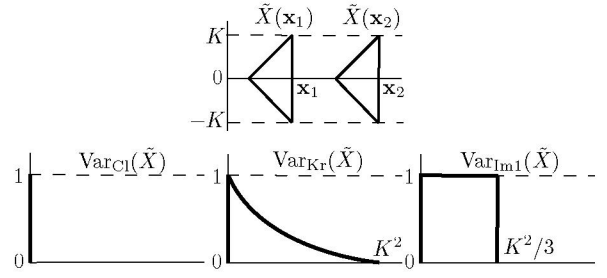

Fig. 2. The three different definitions of variance explained in the text are summarized here for the data in the example 3.1. The classical variance of the residual of this model is 0, which is not adequate. Kruse's et al variance, when extended to frv, produces a valid result but it is not intuitive how to define dominance relations between them. The first order variance produces an interval, which is more convenient from a numerical point of view.

and, for every $\alpha \in (0,1]$, $P^*_{\tilde{D}_\alpha}$ is Dempster's upper probability of the random set $\tilde{D}_\alpha$.

The FMSE of a fuzzy residual is, as we will see in section III-D, easy to implement in certain multicriteria search techniques. Before explaining how, let us detail first how to estimate the FMSE of a model from a sample of residuals.

C. Application of the profile method to compute the bounds of the squared error from a sample

The same procedure described in [13] to estimate the variance of a frv can be applied to estimate the FMSE from a sample of residuals. For each of the membership functions $\tilde{D}(\mathbf{x}_1), \tilde{D}(\mathbf{x}_2), \ldots, \tilde{D}(\mathbf{x}_n)$, we compute the fuzzy set $\tilde{D}^2$, whose α-cuts are $\tilde{D}^2_\alpha = \{x^2 | x \in \tilde{D}_\alpha\}$. Since the function x^2 is not locally monotonic, to evaluate the image of a fuzzy set we must divide the area under the membership functions in zones separated by the changes in the slope of this function. This is graphically illustrated in Figure 3. If the membership of $\tilde{D}$ does not cut the line $x = 0$, the number of vertices is preserved. Otherwise, the left part of the profile is replaced by a vertical segment, and the new right profile is the maximum of the squares of the former left and right parts. After we have computed all of the $\tilde{D}^2(\mathbf{x}_i)$, we compute the FMSE of each one of them. Let M_i^- be the left profile of $\tilde{D}^2(\mathbf{x}_i)$, and M_i^+ the right one. Then,

$$\text{FMSE}_i = \left[\int_0^1 M_i^- d\alpha, \int_0^1 M_i^+ d\alpha \right] \quad (7)$$

and the FMSE of the whole model is

$$\text{FMSE} = \frac{1}{n} \left(\text{FMSE}_1 \oplus \ldots \oplus \text{FMSE}_n \right) \quad (8)$$

D. Use of the FMSE in the NSGA-2 algorithm

In this point, we have all the tools that we need to implement an optimization algorithm that searches for the model that best fits our fuzzy data. We have decided to adopt the NSGA-2 genetic algorithm [4], and to modify it for admitting fuzzy data and an interval-valued fitness function. We need to provide:

1) an alternate definition of the precedence operator, that will be used to assign each individual its non-domination rank, and
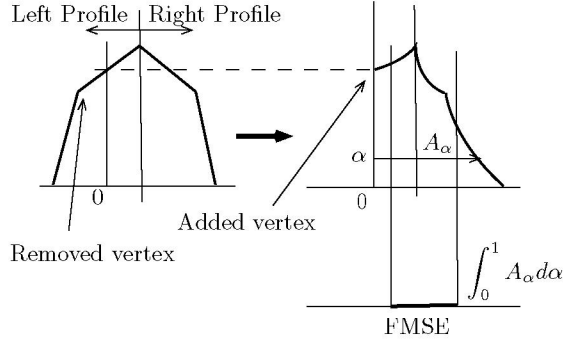
Fig. 3. The same extension to nonmonotonic functions of the profile method that is used to compute the sample variance of a frv can be applied to obtain its FMSE.

2) a definition of the crowding distance used to preserve the diversity.

We propose to use the bounds of the FMSE as two different numerical criteria. This solves the definition of the crowding distance, which is not different from the standard case. The dominance relation is not so immediate, though. Given two models f_1 and f_2, with FMSEs $[a^-, a^+]$ and $[b^-, b^+]$, the dominance is defined as follows: $f_1 \prec f_2$ if $a^+ < b^-$, $f_2 \prec f_1$ if $b^+ < a^-$, and neither of them dominates the other, otherwise. It is intuitively clear that we can only say that one model is better than other when their FMSE do not intersect, therefore we say that an individual dominates the other if the upper bound of the first is lower that the lower bound of the second.

IV. PRACTICAL APPLICATION: CAUSAL MODELING IN MARKETING

In this section we briefly detail how the fuzzy set-based representation of data and its associated imprecise probabilities-based fitness function can be implemented in a practical problem of casual modeling in marketing. In this respect, we focus this section on the modeling estimation techniques by providing a knowledge extraction method that provides more quantity of qualitative information than preceding estimation techniques used in this field [14].

A. Acquisition of data

Data are obtained by means of a questionnaire. Specifically, in Table I we show a hypothetical example of the set of items that could have been used for measuring each one, while Table II shows an example of data available for this problem.

As we have mentioned in the introduction, To work with this unusual kind of data, one could think on reducing the items of a specific variable to a single value, but we have adopted a more sophisticated process that allows us to take profit from the original format without any pre-processing stage: the consideration of fuzzy numbers to describe each variable, as described in Section II.

B. Definition of the Genetic Fuzzy System

Once fixed the linguistic variables, a genetic fuzzy system is proposed in this section to automatically extract the knowledge

Fashion consciousness	
f_1:	Fashion is an important means of self-expression
f_2:	I'm usually the first among my friends to learn about a new brand or product
Conservatism	
c_1:	I tend to achieve my goals one step at a time
c_2:	I'm the type to deliberate things
c_3:	I gather various information and study well when deciding to buy a specific item
Hedonism	
h_1:	I want to enjoy the present rather than think about the future
h_2:	I like to go out to night-time entertainment spots
h_3:	I want to lead a life with lots of ups and downs

TABLE II

EXAMPLE OF FOUR RESPONSES ABOUT THE ITEMS SHOWN IN TABLE I

Fashion consciousness		Conservatism			Hedonism		
f_1	f_2	c_1	c_2	c_3	h_1	h_2	h_3
2	3	7	6	5	2	3	3
6	6	2	3	3	8	7	7
8	7	2	1	2	7	8	9
5	5	2	2	2	7	7	7

existing in the considered fuzzy data. The obtained model should not be only accurate enough but also be easily legible, therefore we consider a multiobjective genetic fuzzy system, whose main components are described in the following sections.

1) Fuzzy Rule Structure: We opt by a compact description based on the disjunctive normal form (DNF) [16]:

$$\text{IF } X_1 \text{ is } \widehat{A_1} \text{ and } \ldots \text{ and } X_n \text{ is } \widehat{A_n} \text{ THEN } Y \text{ is } B$$

where each input variable X_i takes as a value a set of linguistic terms $\widehat{A_i} = \{A_{i1} \vee \ldots \vee A_{il_i}\}$, whose members are joined by a disjunctive (T-conorm) operator, whilst the output variable remains a usual linguistic variable with a single label associated. for instance, a fuzzy rule of the model given as example could be as follows:

IF *FashionConsciousness* is A_1 and *Conservatism* is A_2
THEN *Hedonism* is B.

2) Coding scheme: Each individual of the population represents a set of fuzzy rules (i.e., Pittsburgh style). Each chromosome consists of the concatenation of a number of rules. The chromosome size is variable-length. Each rule (part of the chromosome) is encoded by a binary string for the antecedent part and an integer coding scheme for the consequent part. The antecedent part has a size equal to the sum of the number of linguistic terms used in each input variable. The allele '1' means that the corresponding linguistic term is used in the corresponding variable. The consequent part has a size equal to the number of output variables. In that part, each gene contains the index of the linguistic term used for the corresponding output variable.

For example, assuming we have three linguistic terms (S, M, and L) for each input/output variable, the fuzzy rule [IF X_1 is S and X_2 is {M or L} THEN Y is M] is encoded as [100|011||2]. Therefore, a chromosome would be the concatenation of a

number of these fuzzy rules, e.g., [100|011||2 010|111||1 001|101||3] for a set of three rules.

3) Objective Functions: In addition to the two objectives originating in the fuzzy approximation error and discussed in section III-D, we also add a third objective that intends to assess the linguistic complexity of the generated fuzzy rule set. We measure the number of rules of the fuzzy system $\mathcal{F}$ as $C_1(\mathcal{F})$. However, since each DNF-type fuzzy rule has also a complexity degree itself, we should also consider this aspect. Then, let $C_2(\mathcal{F}) = \sum_{R_r \in \mathcal{F}} \prod_{i=1}^{n} l_{ri}$ be the complexity of the fuzzy system $\mathcal{F}$, with l_{ri} being the number of linguistic terms used in the ith input variable of the rth DNF-type fuzzy rule. The total number of available linguistic terms is computed when an input variable is not considered (i.e. "don't care"). The joint objective is the product of both complexities.

4) Evolutionary Scheme: A generational approach with the multiobjective NSGA-II replacement strategy [4] is considered. Binary tournament selection based on the crowding distance in the objective function space is used.

5) Genetic Operators: The *crossover* operator randomly chooses a cross point between two fuzzy rules at each chromosome and exchanges the right string of them. Therefore, the crossover only exchanges complete rules, but it does not create new ones since it respects rule boundaries on chromosomes representing the individual rule base. In the case that inconsistent rules appear after crossover, the ones whose antecedent is logically subsumed by the antecedent of a more general rule are removed. Redundant rules are also removed.

The *mutation* operator randomly selects an input or output variable of a specific rule. If an input variable is selected, one of the three following possibilities is applied: *expansion*, which flips to '1' a gene of the selected variable; *contraction*, which flips to '0' a gene of the selected variable; or *shift*, which flips to '0' a gene of the variable and flips to '1' the gene immediately before or after it. The selection of one of these mechanisms is made randomly among the available choices (e.g., contraction can not be applied if only a gene of the selected variable has the allele '1'). If an output variable is selected, the mutation operator simply increases or decreases the integer value. In the same way, specific rules appeared after mutation are subsumed by the most general ones and redundant rules are removed.

V. EXPERIMENTAL RESULTS

The consumer behavior model we have used for the experimentation is based on analyzing the consumer's *flow state* in interactive computer-mediated environments. Data have been obtained from the survey used in [17] to test a conceptual model previously presented by the same authors. We have adapted the original structural model proposed in that work by removing the least significant latent variable in each second-order variable. According to the partition performed by the authors, training data is composed by 1,154 examples (consumers' responses) and test data by 500 examples. As an example, we focus the analysis on a specific relationship among the six relationships with a total of 12 variables available in the data set.
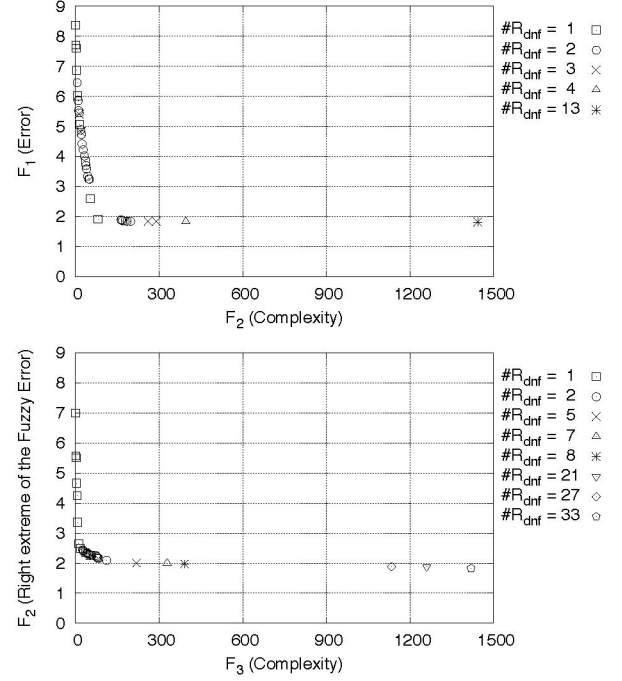

Fig. 4. Crisp (upper part) and projection over the second (upper bound of the fuzzy error) and third (complexity) criteria of the fuzzy (lower part) joint Pareto-front

We have run 10 times the proposed genetic fuzzy system, and compared its performance to that of a crisp error-based multiobjective approach in [18]. The resulting joint Pareto-fronts are displayed in Figure 4. Notice that in the fuzzy case, the Pareto is projected over the second (upper bound of the fuzzy error) and third (complexity) criteria. Although this shows the worst case of the fuzzy error, it allows us to compare the accuracy degree of crisp and fuzzy versions. Observe that there are not important differences between them. But the behavior of both is significantly different when we select a different measure of accuracy: the average of the differences between the output of these models and *all* the items in every output variable. Moreover, it is remarked that the result of this average is not a criteria being minimized by the genetic algorithm. If it were, the instances with a higher dispersion would contribute to the global error more than the instances with low variance, thus we would be giving more importance to the questionnaires with conflicting items. Either the crisp and the fuzzy implementations of the fitness function were meant to compare the output of the model with a characteristic value of the set of items comprising the output of each instance, and the measurements of accuracy in the Pareto front reflect this design. But, given two models with similar instance-wise error, we definitely prefer the model with the best item-wise error, since it describes better the relationship between the input and disperse sets of output items.

To prove this point, we include in Figures 5 and 6 the convergence plots and two boxplots of the item-wise error in both the fuzzy and crisp versions of the algorithm. Observe that the maximum value of test error in the fuzzy fitness is almost always better than the minimum value of the scalar fitness,
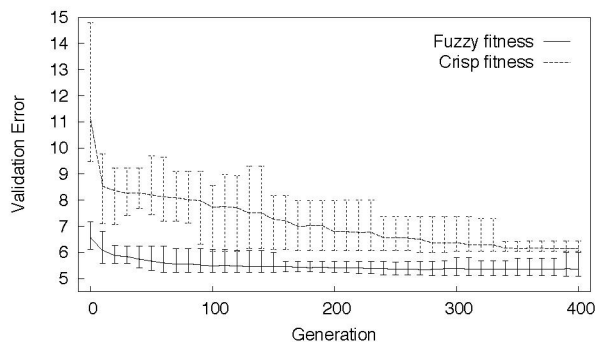
Fig. 5. Minimum, mean and maximum values of the validation error for the crisp problem (dash lines) and the fuzzy problem (solid lines)
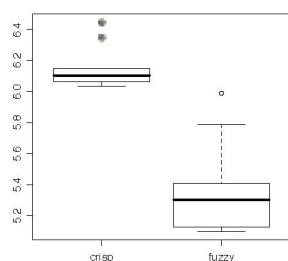

Fig. 6. Boxplot showing the differences between the item-wise errors of the crisp and fuzzy models after 400 generations and 10 runs of the experiments.

and the boxplot evinces a statistically significant difference.

VI. CONCLUDING REMARKS

In this paper we have put into practice a recent semantic interpretation of fuzzy sets. We have considered that a fuzzy set is equivalent to a nested family of confidence intervals for the value of an unknown parameter. By doing this, we have converted data comprising multi-item examples into fuzzy data. A knowledge extraction methodology has also been proposed. That methodology is able to obtain a model, based on fuzzy association rules, from the vague data that was originated in the multi-item examples. We have shown, with the help of a practical problem, that the models obtained minimizing the bounds of the squared error between two fuzzy random variables are more robust than "state of the art" genetic fuzzy models, and are able to capture the dependence between imprecise data without the need of aggregating them or removing their fuzziness.

REFERENCES

[1] J. Casillas, F. Martínez-López, and F. Martínez, "Fuzzy association rules for estimating consumer behaviour models and their application to explaining trust in internet shopping," *Fuzzy Economic Review*, vol. IX, no. 2, pp. 3–26, 2004.

[2] I. Couso, S. Montes, and P. Gil, "The necessity of the strong alpha-cuts of a fuzzy set," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 2, pp. 249–262, 2001.

[3] L. Sánchez and I. Couso, "Advocating the use of imprecisely observed data in genetic fuzzy systems," admitted for publication in IEEE Transactions on Fuzzy Systems.

[4] K. Deb, A. Pratap, S. Agarwal, and T. Meyarevian, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[5] I. R. Goodman, "Fuzzy sets as equivalence classes of possibility random sets," in *Fuzzy Sets and Possibility Theory: Recent Developments*, R. R. Yager, Ed. Pergamon, Oxford, 1982.

[6] V. Kreinovich, "Range estimation is np-hard for ϵ^2 accuracy and feasible for $\epsilon^{2-\delta}$." *Reliable Computing*, vol. 8, no. 6, pp. 481–491, 2002.

[7] W. Dong and H. C. Shah, "Vertex method for computing functions of fuzzy variables," *Fuzzy Sets Syst.*, vol. 24, no. 1, pp. 65–78, 1987.

[8] D. Dubois, H. Fargier, and J. Fortin, "A generalized vertex method for computing with fuzzy intervals ," in *Proc. of the International Conference on Fuzzy Systems , Budapest, Hungary*. IEEE, 25-29 juillet 2004, pp. 541–546.

[9] I. Couso, S. Montes, and L. Sanchez, "Varianza de una variable aleatoria difusa. estudio de distintas definiciones," in *Proc XII Congreso Espanol sobre Tecnologias y Logica Fuzzy*, 2004.

[10] R. Körner, "On the variance of fuzzy random variables," *Fuzzy Sets and Systems*, vol. 92, pp. 83–93, 1997.

[11] M. A. Lubiano, M. A. Gil, M. Lopez-Diaz, and M. T. Lopez, "The λ-mean squared dispersion associated with a fuzzy random variable," *Fuzzy Sets Syst.*, vol. 111, no. 3, pp. 307–317, 2000.

[12] K. D. Meyer and R. Krüse, *Statistics with vague data*. D. Reidel Publishing Company, 1987.

[13] D. Dubois, H. Fargier, and J. Fortin, "The empirical variance of a set of fuzzy intervals ," in *Proc. of the 2005 IEEE International Conference on Fuzzy Systems , Reno, Nevada*. IEEE, 2005, pp. 885–890.

[14] G. V. Bruggen and B. Wierenga, "Broadening the perspective on marketing decision models," *International Journal of Research in Marketing*, vol. 17, pp. 159–168, 2000.

[15] S. MacLean and K. Gray, "Structural equation modelling in market research," *Journal of the Australian Market Research Society*, vol. 6, pp. 17–32, 1998.

[16] A. González and R. Pérez, "Completeness and consistency conditions for learning fuzzy rules," *Fuzzy Sets and Systems*, vol. 96, no. 1, pp. 37–51, 1998.

[17] Y. Novak, D. Hoffman, and Y. Yung, "Measuring the customer experience in online environments: a structural modelling approach," *Marketing Science*, vol. 19, no. 1, pp. 22–42, 2000.

[18] J. Casillas, O. Delgado, and F. Martínez-López, "Predictive knowledge discovery by multiobjective genetic fuzzy systems for estimating consumer behavior models," in *Proceedings of the 4th International Conference in Fuzzy Logic and Technology*, Barcelona, Spain, 2005, pp. 272–278.