Knowledge Extraction from Fuzzy Data for Estimating Consumer Behavior Models

Jorge Casillas and Luciano Sánchez

Abstract— For certain problems of casual modeling in marketing, the information is obtained by means of questionnaires. When these questionnaires include more than one item for each observable variable, the value of this variable can not be assigned a number, but a potentially scattered set of values.

In this paper, we propose to represent the information contained in this set of values by means of a fuzzy number. A novel 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 a central tendency measure of the value of the variable. A genetic learning algorithm, able to extract association fuzzy rules from this data, is also proposed. The accuracy of the model will be expressed by means of a fuzzy-valued function. We propose to jointly minimize this function and the complexity of the rule based model with multicriteria genetic algorithms, that in turn will depend on a fuzzy ranking-based ordering of individuals.

I. INTRODUCTION

In previous works [1], we have stated that, when a variable is composed of a set of parameters (items), the data is uncertain, because each one of the items provide only partial information to describe the variable, and these informations may be in conflict. Therefore, the use of association fuzzy rules to extract knowledge from casual models in marketing has some inherent advantages:

- Since the models have to relate uncertain variables, the fuzzy logic framework is *a priori* well suited to express the relationships between them.
- The obtained fuzzy models can be linguistically interpreted, and this fact provides the researcher with an additional insight into the data.

Regarding the first benefit, the conversion of a set of items into a compound value that can be fed to the model has been solved in different ways. The classical solution [2] consists in preprocessing these sets of values, then 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 the model should know not only about the characteristic values of the variable, but also about the degree of imprecision with which these values are known. More recently [1], a mechanism that assigns a degree of truth to any fuzzy assert about the value of the variable, given the mentioned set of values, has also been proposed. This

second approach is better, because it allows to take profit from the original format without any preprocessing stage, but it can be further improved. In this paper, we apply a novel interpretation of the semantics of a fuzzy set [3], that allows us to define a mapping between sets of items and fuzzy numbers. Therefore, we will model each set of items by means of a fuzzy number, either for the input and the output variables.

The definition of the learning problem that arises involves optimizing a fuzzy valued function, because the accuracy of the model is measured by the squared difference between the fuzzy representation of the output variables, and the images of the input values provided by the candidate model, which are also fuzzy [4]. As done in previous works, we will jointly optimize the precision of the model and its complexity, by mean of multicriteria genetic algorithms. The purpose of the learning is to find a model both accurate and linguistically understandable.

The paper is organized as follows. Section II briefly describes the dealt problem based on consumer behavior models. Section III introduces the fuzzy representation issues, and Section IV details the proposed methodology. Section V shows some obtained experimental results. Finally, Section VI concludes.

II. CAUSAL MODELING IN MARKETING: A KIND OF FUZZY DATA

Marketing academics and practitioners have emphasized the need for knowing and explaining the consumer's behavior patters in a manner increasingly efficient. This is mainly due to firms focused on final markets are immersed in highly competitive systems in which it is needed that their decision processes to be as correct as possible. In this sense, models of consumer behavior are considered as a specific case of *marketing management support system*, and throughout the time have demonstrated to be a source of transcendental relevance for the development of marketing science [5].

Notwithstanding, current models of consumer behavior do not seem to cover all the necessities that it should supposedly satisfy a model which aims to aid on the marketing decision making. Thus, as the main problem that actually face firms oriented to consumer markets is not the availability of information (data), but the possession of appropriate levels of knowledge to take the right decisions, the use of avant-garde knowledge discovery techniques able to exploit it may represent an essential source of competitive advantage.

In this respect, we focus our paper on the modeling estimation techniques by providing a knowledge extraction

J. Casillas is with the Dept. Computer Science and Artificial Intelligence, University of Granada, E-18071, Spain (email: casillas@decsai.ugr.es)

L. Sánchez is with the Dept. Computer Science, Campus de Viesques, University of Oviedo, E-33071, Spain (email: luciano@uniovi.es)

method that provides more quantity of qualitative information than preceding estimation techniques used in this field [5]. Before presenting the proposal in Section IV, the following subsections introduce the kind of data and information which are available in causal modeling.

A. Data Gathering

First step is to collect the data related to the variables defining the theoretical model of consumer behavior proposed. In this sense, as it has been traditionally done in marketing, data are obtained by means of a questionnaire in a similar way to the models estimated by structural equation modeling. Thus, at first, attention should be paid to how consumer behavior modelers face and develop the measurement process of variables which complex behavioral models contain.

It can be said that measuring streams for these latent variables in consumer modeling can be classified into two groups depending on whether they defend that these constructs can or cannot be perfectly measured by means of observed variables (indicators)—i.e., the existence or not of a one-to-one correspondence between a construct and its measurement. Certainly, though consumer behavior modelers tended to make use in the beginning of what was known as the *operational definition philosophy*, a more convenient and reasonable position is that based on the *partial interpretation philosophy* which distinguished between unobserved (constructs) and observed (indicators) variables. This later approach of measurement, being currently predominant in the marketing modeling discipline, poses to jointly consider multiple indicators—imperfect when considered individually, though reliable when considered altogether—of the subjacent construct to obtain valid measures.

For instance, we can consider the measurement model depicted in Figure 1, compounded by three construct or latent variables (depicted by circles), two exogenous (*fashion consciousness* and *conservatism*), and one endogenous (*hedonism*). Since these latent variables are unmeasurable, we indirectly measure them by means of observable variables (items), depicted by rectangles in the figure.
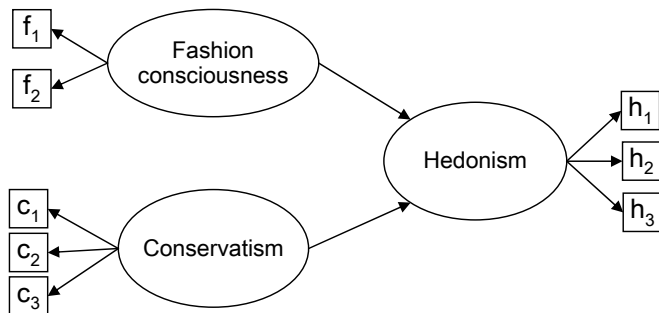

Fig. 1. Example of a simple measurement (structural) model (extracted from [6])

Likewise, with respect to the measurement scales, imagine that the constructs have been measured by means of several nine-points Likert scales ranging from 1: strongly disagree to 9: strongly agree. 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.

TABLE I

EXAMPLE OF A QUESTIONNAIRE ASSOCIATED TO THE MEASUREMENT MODEL SHOWN IN FIGURE 1 (EXTRACTED FROM [6])

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

B. Data Processing

To work with this unusual kind of data, one could think on adapting the collected data to a scheme easily tractable by classical learning method. However, the approach should be aware of the special features of the available data (with several items or indicators to describe a specific variable) when adapting the observed variables to a learning method. An intuitive approach could directly reduce the items of a specific variable to a single value (e.g., by arithmetic mean) [2]. Another possibility would be to expand any multi-item example (the result of a questionnaire filled by a consumer) to several single-item examples and subsequently reduce the data size with some instance selection process.

The problem of these approaches is that the data are transformed, so relevant information may be lost or strained. We propose 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. It is described in Section III.

C. Prior Information on the Fuzzy System from the Structural Model

Prior to applying the genetic fuzzy systems to automatically design the fuzzy model, we fix a number of components: the set of variables to be modeled, the transformation of the scales used for measuring such variables into fuzzy semantics, and the fuzzy rule structure

(relations among constructs). As mentioned, the expert is able to provide its knowledge about the problem by a measurement structural model like shown in Figure 1 (of course, a real problem would work with a more complex model). From this information, we can deduce the variables and the direction (in terms of antecedents and consequents) of the relationships existing among them. Therefore, we can easily fix the input and output variable of the analyzed relationship. For example, from the measurement model of Figure 1, the fuzzy rule structure have the following form:

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

With respect to the fuzzy semantic used for each variable, it is also possible to fix it according to expert knowledge. Indeed, when she/he build the questionnaire in order to collect data, she/he fixes the kind of scale and precision (number of points) used to measure each variable. From this information it is possible to define a fuzzy semantic. At this point, several marketing scale types can be used for its measurement. With the aim of simplifying the problem, in this paper we focus on interval scale (i.e., Likert differential semantic or rating scale), which is one of the most commonly used in marketing.

We suggest to transform these scales to Ruspini's strong fuzzy semantics with uniform density of the fuzzy membership functions to statistically unbias the significance of every linguistic term. Thus, we define the membership function shapes such as, given the set $S = \{min, \ldots, max\}$ defining an interval variable, they hold the following condition:

$$\sum_{k \in S} \mu_{A_i}(k) = \frac{max - min}{l}, \quad \forall A_i \in \mathcal{A}, \qquad (1)$$

with l being the number of linguistic terms and $\mathcal{A} = \{A_1, \ldots, A_l\}$ the set of them.

Figure 2 shows an example based on the transformation of a nine-point rating scale (a typical marketing scale used to measure the observed variables related to a construct) into a fuzzy semantic with the three linguistic terms *Low*, *Medium*, and *High*.
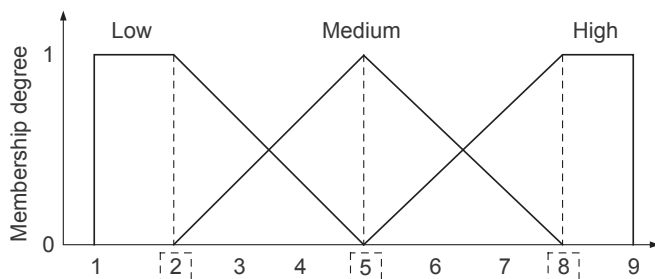

Fig. 2. Transformation of a nine-point rating scale into a three-linguistic-terms fuzzy semantic

At this stage, one could think about using some mechanism to automatically generate fuzzy partitions from data [7], or design a tuning method to adapt uniformly initialized fuzzy

semantics [8], or both of them. However, in the analyzed problem we are not (more accurately, the expert is not) interested on generating fuzzy semantics that accurately cover data. It is because of the faithfully way to interpret the semantic considered by each consumer that filled the form is the uniform one. If we apply any automatic process to generate/tune fuzzy membership functions, we are adapting to the context, i.e., the answers of the consumer, but not to the meaning of the variables. Therefore, in this problem the knowledge discovery process is focused on the relationship among the variables (fuzzy rule surface structures).

III. REPRESENTATION ISSUES

The input and output data comprises multi-item values (as discussed in Section II). In previous works, different aggregation functions were proposed [1]. These functions ultimately lead to the assignment of certain truth value to the assert "the value of the item is V" for certain label V of a suitable linguistic variable. Here we will be taking a different path assuming that there exists a true value for the multi-item variable, but also that this value is unknown and we can know, at the best, a set that contains it.

A. Semantics of a Fuzzy Set

The amplitude of this set can be understood as a tolerance in the measuring of this last value. According to this, a fuzzy set will represent a set of tolerances, each one of them is assigned a confidence degree, being the lower degree the narrower tolerance. Following the semantic interpretation of a fuzzy set in [3], which in turn is related to that proposed in [9], a nested family of confidence sets can be represented by a fuzzy set, whose α-cuts are confidence intervals with degree $1 - \alpha$ (see an example in Figure 3.)
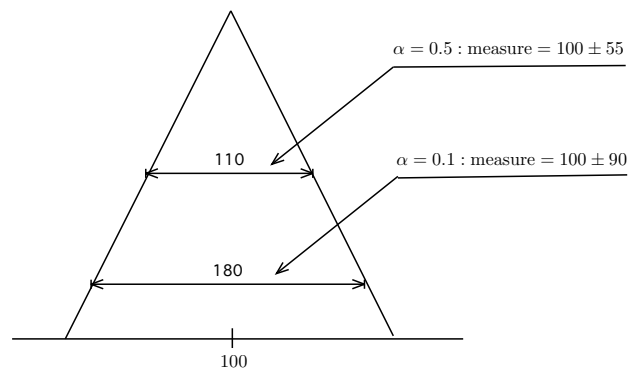

Fig. 3. Fuzzy sets can be assimilated to sets of confidence intervals for an unknown parameter of a distribution. An α-cut of this set is the smallest interval that contains the true parameter with probability greater or equal than $1 - \alpha$

B. Numerical Procedure of Fuzzification

The numerical method for translating a set of numbers into a fuzzy set will be made clear with the example that follows. Let us suppose that a latent variable X has associated the items valued

$$X = \{2, 1, 3, 3, 2, 2, 4\}. \qquad (2)$$

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. To simplify the calculations, let us assume that the sample was drawn from a normal population. Then, the confidence intervals for the mean of the population have the form

$$C(\alpha) = 2.429 \pm 0.9759 \cdot \text{qt}_6 \left(1 - \frac{1-\alpha}{2}\right), \quad (3)$$

where qt_6 is the quantile function for the t distribution, with 6 degrees of freedom, and therefore, the α-cuts of the fuzzy set $\widetilde{X}$ that represents the value of the variable are the intervals

$$\widetilde{X}_\alpha = 2.429 \pm 0.9759 \cdot \text{qt}_6 \left(1 - \frac{\alpha}{2}\right). \quad (4)$$

A graphical representation of the membership function of $\widetilde{X}$ is shown in Figure 4. Observe that we can approximate it by a triangular membership function without incurring large errors.

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 and, in particular, it does for the problem that will be described in Section V. Otherwise, or when the data contains many outliers, building the membership function from the quantiles of the bootstrap distribution of the median should be a sensible choice.
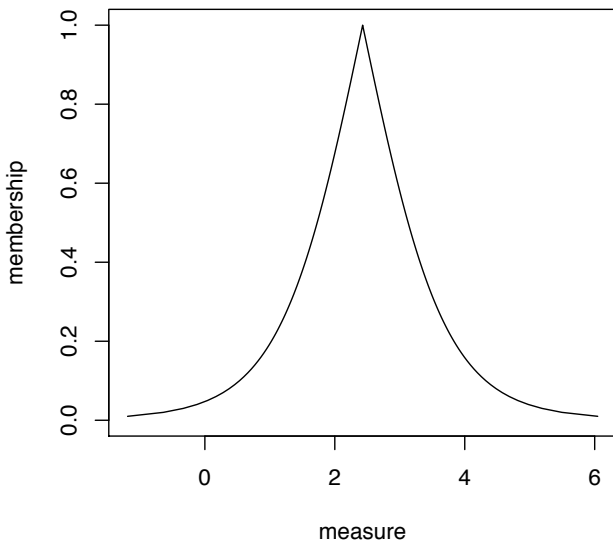

Fig. 4. Membership function of the set $\widetilde{X}$ that represents the sample X in Section III-B. This membership can be approximated by a triangular membership function with reasonable accuracy

IV. Proposed Genetic Fuzzy System

Once fixed the linguistic variables that properly represent the tackled available information, a genetic fuzzy system is proposed in this section to automatically extract the knowledge existing in the considered data. The algorithm has been designed to perform predictive induction [10], that looks for generating models that describe with the highest reliability the data set that represent the analyzed system. However, the obtained model should not be only accurate enough but also be easily legible in order to be able to linguistically describe the real system. As it is known, accuracy and interpretability are two contradictory properties. To address that, we consider a multiobjective genetic fuzzy system thanks to their good behavior to deal with multiple, contradictory objectives. Next subsections describe the main components of the proposed method.

A. Fuzzy Rule Structure

In data mining is crucial to use a learning process with a high degree of interpretability. Therefore, we opt by a compact description based on the disjunctive normal form (DNF) [11]. This kind of fuzzy rule structure has the following form:

IF X_1 is $\widehat{A_1}$ and ... and X_n is $\widehat{A_n}$ **THEN** Y 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. We use the *bounded sum* ($min\{1, a + b\}$) as T-conorm. The structure is a natural support to allow the absence of some input variables in each rule (simply making $\widehat{A_i}$ to be the whole set of linguistic terms available).

B. 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 number of rules is not fixed *a priori* so, 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.

C. Objective Functions

We consider two objective functions to assess the quality of the generated fuzzy systems, the former (fuzzy approximation error) to improve the accuracy and the latter (linguistic complexity) to improve the interpretability.

1) Fuzzy Approximation Error: We will use the mean squared error (MSE) to measure the differences between the output of the model and the desired values. Though, we have stated that we cannot know the precise input neither output data, but fuzzy sets that describe them (see Section III). Therefore, we can not compute a number that measures the error of a candidate model over our train data, but we can provide upper and lower bounds of it. Following [4], let our model $\mathcal{F}$ be a fuzzy rule base that maps an input vector $\mathbf{x}$ to an output value $\mathcal{F}(\mathbf{x})$. If all we can say about $\mathbf{x}$ is that it is contained in the fuzzy set $\widetilde{X}$, then $\mathcal{F}(\mathbf{x})$ is contained in the fuzzy set $\mathcal{F}(\widetilde{X})$, defined by the family of α-cuts

$$\mathcal{F}(\widetilde{X})_\alpha = \{\mathcal{F}(\mathbf{x}) \mid \mathbf{x} \in \widetilde{X}_\alpha\} \tag{5}$$

and, consequently, the MSE is

$$\widetilde{E}(\mathcal{F}) = \frac{1}{N} \bigoplus_{e=1}^{N} \mathrm{SQ}\left(\mathcal{F}(\widetilde{X}) \ominus \widetilde{Y}\right), \tag{6}$$

where $[\mathrm{SQ}(X)]_\alpha = \{x^2 \lfloor x \in X_\alpha\}$.

Observe that, being $\widetilde{E}(\mathcal{F})$ a fuzzy set, we will need to minimize an imprecisely known function in order to find the best model. This problem can be solved with the help of a fuzzy ranking, that defines total order over the set of errors. In this paper, we decided to use the centroid-based fuzzy ranking proposed in [12]. It basically consists on obtaining the centroid point (x_i, y_i) of each fuzzy number and computing the distance index to the original point, i.e.

$$F_1(\widetilde{E}(\mathcal{F})) = \sqrt{\|x_i\|^2 + \|y_i\|^2}. \tag{7}$$

The values x_i and y_i must be normalized before computing the distance. Since $y_i \in [0, 0.5]$, $\|y_i\| = 2y_i$. The lower the value, the better the corresponding model. In our case, this ranking makes sense because, when the models being compared have non-overlapping fuzzy errors, it produces the fuzzy number that corresponds to the lowest error and, when the errors overlap, the most specific one is chosen.

2) Linguistic Complexity: This second objective intends to assess the linguistic complexity of the generated fuzzy rule set. Firstly, it is clear that higher number of rules, higher complexity. Therefore, 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").

Therefore, the joint objective (to be minimized) is the combination of both complexities as follows:

$$F_2(S) = C_1(\mathcal{F}) \cdot C_2(\mathcal{F}) \tag{8}$$

We opt for this combined measurement instead considering two independent objectives because both are deeply related and assess the same concept (complexity of the system).

D. Evolutionary Scheme

A generational approach with the multiobjective NSGA-II replacement strategy [13] is considered. Binary tournament selection based on the crowding distance in the objective function space is used.

E. 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 AND INTERPRETATION

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 [14] to test a conceptual model previously presented by the same authors. We have adapted the original structural model proposed in [14] by removing the lest significant latent variable in each second-order variable. To illustrate the flow state concept, it is achieved when the consumer is so deeply involved in the process of navigation on the Web that "nothing else seems to matter" [14].

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. The four constructs used as input variables of the system (*interaction speed*, *skill*, *arousal*, and *telepresence*) are considered as primary antecedents of the consumer's flow state.

Interaction speed collects the opinion of the individual with respect to the answering speed of the environment (quality of the interaction) when she/he is surfing. *Skill* is the ability of the individual to use the Web. *Arousal* stands for the perception of the consumer about the stimulus level that represents surfing on the Web. *Telepresence* is a recent concept associated to computer-based environments. It relates to the state the consumer reaches when surfing on the Web that make her/him feel belonged to the virtual environment where she/he is interacting. It is been hypothesized that these four latent variables are positively related to the central construct (flow).

We have run 10 times the proposed genetic fuzzy system. The fuzzy semantic shown in Figure 2 is considered for all variables. The resulting joint Pareto-front and average convergence plot are depicted on Figures 5 and 6, respectively.
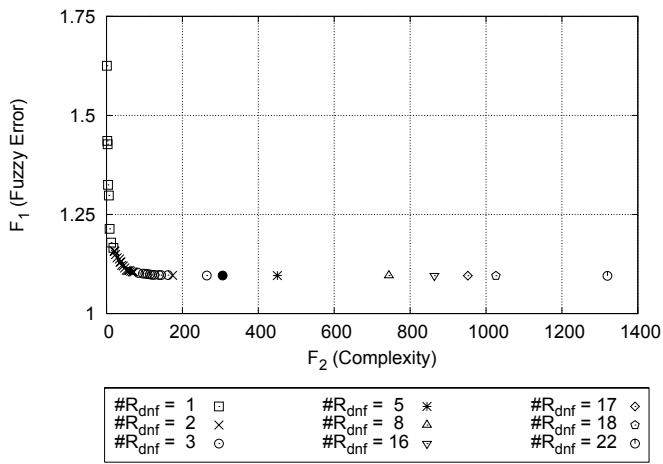

Fig. 5. Joint Pareto-front. Each symbol type represents fuzzy models with a specific number of DNF-type fuzzy rules—$C_1(\mathcal{F})$. The solid symbol ($\bullet$) corresponds to the solution shown in Table III
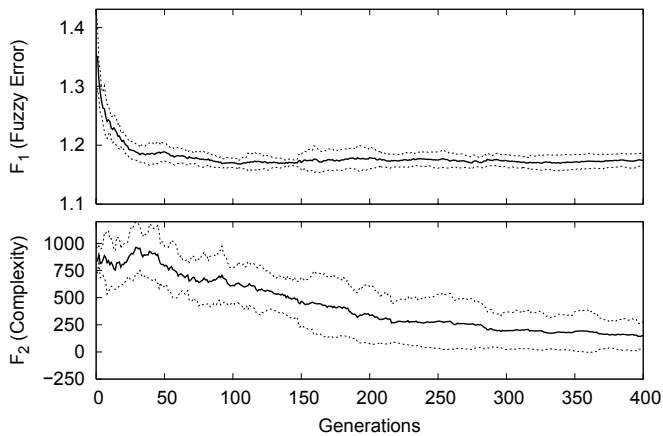

Fig. 6. Convergence plot (solid lines) of mean error (top) and mean complexity (bottom) of the Pareto set. Dotted lines show $\bar{x} \pm \sigma$

From Figure 5, we can observe that the algorithm is able to obtain a wide collection of solutions with different interpretability-accuracy tradeoffs. It is interesting to see how
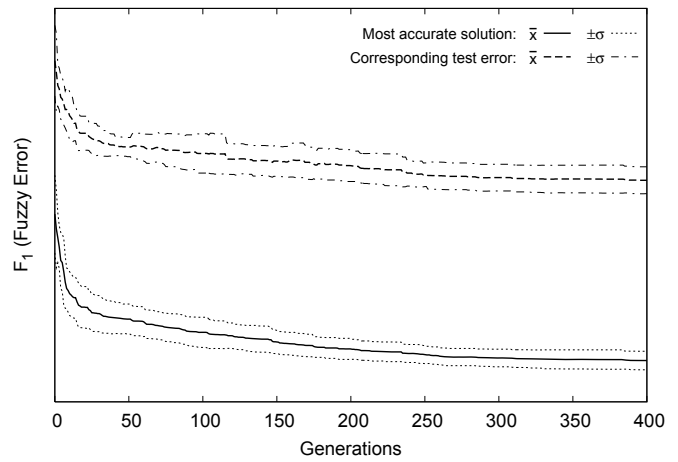

Fig. 7. Convergence plot of the most accurate solution (solid line) and its corresponding test error (dashed line) obtained by the proposed algorithm. Dotted lines show $\bar{x} \pm \sigma$
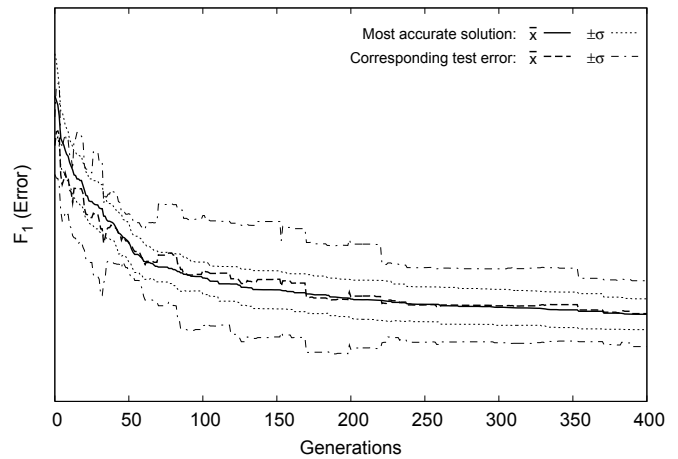

Fig. 8. Convergence plot of the most accurate solution (solid line) and its corresponding test error (dashed line) obtained by the algorithm proposed in [1]. Dotted lines show $\bar{x} \pm \sigma$

there is a threshold, around 5 DNF-type fuzzy rules, where improving the accuracy is only possible at the expense of significantly increasing the number of rules. Furthermore, the fact of obtaining accurate fuzzy models with a low number of DNF-type rules but a high number of equivalent Mamdani rules shows the ability of the algorithm to obtain rules as abstract (general) as possible. (We have to say that the inference process was designed to return a value outside the domain when no rules are matched in order to avoid over-abstraction.) Figure 6 shows the good convergence and robust behavior of the algorithm, where the complexity is gradually decreased without worsening the accuracy and the standard deviation of the 10 runs is low.

Moreover, Figure 7 shows the mean and standard deviation values of the most accurate solution obtained at each generation and its corresponding error over the test data set. We can observe the robust behavior of the proposed algorithm and the lack of overfitting during the optimization process. It is due to the fact of considering a fitness based on fuzzy error.

TABLE III

AN OBTAINED FUZZY RULE SET—*mdm* STANDS FOR *medium*. $F_1(\mathcal{F}) = 1.096140$, $F_2(\mathcal{F}) = 306$

Interaction Speed			Skill			Arousal			Telepresence			Flow		
low	mdm	high	low	mdm	high	low	mdm	high	low	mdm	high	low	mdm	high
										×	×		×	
				×	×	×		×	×				×	
×		×							×		×	×		

This behavior is more clear if we compare with the results (shown in Figure 8) obtained by the method proposed in [1] that does not consider a fuzzy fitness function. In this latter case, the progression of the test error is very unstable and the standard deviation among the 10 runs is higher.

Likewise, Table III shows an example of the fuzzy models available in the Pareto optimal set. According to this example fuzzy model, we can observe that the three first constructs (input variables) seems to exert a poor influence over the consumer's flow state, even with no repercussions depending on the case. However, *telepresence* is, with no doubt, the most relevant construct. As it can be easily seen, it plays a key role to specially determine low and medium levels of flow state and shows a positive (direct) relationship with it. It is also interesting to highlight that, when *telepresence* is low, medium degrees of flow can still be obtained if *skill* and *arousal* are high. Furthermore, the *interaction speed* variable is only significative in combination with low degrees of *telepresence* to obtain a low degree of flow.

To sum up, we can deduce several interesting relationships between the variables that would be useful for the marketing decision makers. On the one hand, it is difficult to obtain a high degree of flow. On the other, it is possible to obtain a medium degree of the consumer's flow by providing an attractive surfing process (good telepresence) with no matter about the experience or the stimulus of the consumer. When the telepresence is low, consumers with highly experienced or stimulated can still get medium degrees of flow. Finally, when the speed of the computer environment and the telepresence take extreme values (low or high), the consumer's flow is low.

VI. Concluding Remarks

The paper has introduced a novel problem—causal modeling in marketing—where knowledge extraction by genetic fuzzy systems can help to generate highly understandable fuzzy models for predictive induction. The problem provides a specific kind of uncertain data set that justifies the use of fuzzy rules. We develop multiobjective optimization to obtain accurate and legible fuzzy models. The proposed knowledge extraction methodology has been appropriately applied to a real-world causal modeling problem that analyzes interactive computer-mediated environments.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] S. Guillaume and B. Charnomordic, "Generating an interpretable family of fuzzy partitions from data," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 3, pp. 324–335, 2004.

[8] J. Casillas, O. Cordón, M. del Jesus, and F. Herrera, "Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 13–29, 2005.

[9] 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.

[10] N. Lavrac, B. Cestnik, D. Gamberger, and P. Flach, "Decision support through subgroup discovery: three case studies and the lessons learned," *Machine Learning*, vol. 57, no. 1-2, pp. 115–143, 2004.

[11] 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.

[12] C.-H. Cheng, "A new approach for ranking fuzzy numbers by distance method," *Fuzzy Sets and Systems*, vol. 95, pp. 307–317, 1998.

[13] 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.

[14] 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.