Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling ☆

Abstract

The main problem currently faced by market-oriented firms is not the availability of information (data), but the possession of appropriate levels of knowledge to take the right decisions. This is common background for firms. In this regard, marketing professionals and scholars highlight the necessity for knowing and explaining consumers' behaviour patterns in an increasingly efficient way. The use of new knowledge discovery methods, able to exploit such data, may represent a relevant source of competitive advantage.

In marketing, the information about most consumer variables of interest is usually obtained by means of questionnaires containing a diversity of items. It is also frequent that marketing modellers make use of unobserved variables to build the consumer models; i.e., abstract variables that need to be measured by means of a set of observed variables or items associated with them. In these cases, the value of a certain unobserved variable cannot be assigned to a number, but to a potentially scattered set of numbers. This fact disables the application of conventional data mining techniques to extract knowledge from them.

In this paper, we present a new approach that is able to deal with this kind of uncertain data by using a multiobjective genetic algorithm to derive fuzzy rules. Specifically, we propose a complete methodology that considers the different stages of knowledge discovery: data collection, data mining, and knowledge interpretation. This methodology is experimented on a consumer modelling application in interactive computer-mediated environments.

Keywords: Consumer behaviour; Fuzzy logic; Genetic algorithms; Knowledge extraction; Machine learning; Marketing

1. Introduction

1.1. Background

Nowadays, it is usual to find numerous academic and professional management-related contributions that show a clear tendency towards defining the current business environment of organizations as hypercompetitive (D'Aveni, 1994). Unlike several decades ago, when companies were mainly concerned with having enough information to guide their decisional processes, the problem faced by organizations has been become increasingly centred, from the middle 1980s till now, on the possession of high quality, thus most valued, and better information about their business framework than their competitors. This is a widely accepted idea that has driven the evolution of the management information systems (MIS), in terms of their design, use and support functions required, from those based on the data, to the ulterior knowledge-based systems whose first exponent were the expert systems (Casey & Murphy, 1994; Sisodia, 1992; Talvinen, 1995; Van Bruggen & Wierenga, 2000).

This question of the improvement of the quality of the systems used to manage the variety of information owned by firms becomes even more important when it is analyzed

within the framework of the marketing function; the business area primarily responsible for managing the relations with consumers, i.e., the firm's target. As Li, Kinman, Duan, and Edwards (2000) note, considering the characteristics related to the competitive environment of firms, marketing strategies directed to markets must be based on an accurate knowledge of the consumers' preferences and behaviour. In this task, the application of suitable marketing management support systems (MkMSS) to the analysis of data plays a notable role (Wierenga & Van Bruggen, 1997, 2000). It is not unusual, therefore, to observe the intensification in the use of knowledge-based MkMSS by firms in recent years (Shim et al., 2002; Wedel, Kamakura, & Böckenholt, 2000).

This evolution of the MkMSS towards systems based on methodologies imported from the artificial intelligence area have tried to fulfil the demands of marketing managers and modellers in terms of working with methods of analysis that are more flexible, powerful and robust, and capable of providing greater and improved information with respect to consumers' behaviour (Lilien, Kotler, & Moorthy, 1992). In this sense, though it is well known that the marketing expert systems were, in line with the MIS framework, the first knowledge-based systems applied to support the marketing managers' decision processes, there have been significant and interesting advances, some of them very recent, such as those based on artificial neural networks, case-based reasoning, clustering, decision trees, or fuzzy systems (Akhter, Hobbs, & Maamar, 2005; Ha & Park, 1998; Li et al., 2000; Wierenga & Van Bruggen, 2000). In any case, regardless of the marketing knowledge-based system we consider, each has one thing in common, the use of knowledge discovery in databases (KDD) methodologies (Fayyad, Piatesky-Shapiro, Smyth, & Uthurusamy, 1996), and hence, data mining (machine learning) paradigms.

KDD implies the development of a process compounded by several stages that allow the conversion of low-level data into high-level knowledge, where the data mining is considered the core stage of such a process (Mitra, 2002). Nevertheless, it is important to be aware of the fact that the application of the data mining stage alone would be insufficient to undertake, with rigor and guarantees of success, a process of KDD (Fayyad & Simoudis, 1995).

It is well known that KDD may offer excellent results when applied to marketing databases in general, as well as to the analysis of the behaviour of consumers in particular, though the development and application of specific KDD methodologies to marketing problems is still incipient (Liao & Chen, 2004). In this regard, we believe that the benefits provided by KDD should not only motivate its use in firms that are clearly interested in improving the efficiency of their MkMSS, thus their marketing decision making, but also in marketing academics. Specifically, in our opinion, the academics' efforts must be focused on two main questions with regard to this: first, an intelligent, oriented and selective increase in the use of the KDD techniques, based on their properties for solving the marketing problem faced by the academics; and second, an active research to adapt generic KDD methodologies to the specific characteristics of the marketing problems to which they are going to be applied. In this sense, the main contribution of this paper focuses on the latter.

1.2. Scope of the paper

In KDD, we can distinguish between two different approaches (Lavrac, Cestnik, Gamberger, & Flach, 2004): *predictive* induction and *descriptive* induction. The difference lies in the main objective pursued and, therefore, the learning method used to attain it. On the one hand, predictive induction looks to generate legible models that describe with the highest reliability the data set that represents the analyzed system. In that case, the goal is to use the model obtained to simulate the system, thus reaching an explanation of its complex behaviour. On the other hand, descriptive induction looks for particular (interesting) patterns of the data set. In that case, we do not achieve a global view of the relationships among variables but we discover a set of rules (different enough among them) that are statistically significant.

This paper focuses on predictive induction to extract useful knowledge guided by theoretical (causal) models used in the discipline of consumer behaviour. In other words, the machine learning stage is driven by a set of relations among variables previously determined by the marketing expert. To do that, we develop a complete KDD methodology, adapted to the kind of causal structures, variables and measurement models usually used in consumer behaviour modelling. Hence, we reflect on and give specific solutions, adapted to the marketing problem we face, to the variety of questions associated with every stage of the KDD process; i.e., pre-processing, machine learning and post-processing. Basically, the benefits we provide with this methodology try to cover the academic as well as the professional fields, though we mainly highlight the interesting qualities of its practical applicability in order to help marketing managers to better predict consumer behaviour. Specifically, association fuzzy rules, with input and output variables previously fixed by the theoretic model of reference, are used. The extraction is performed by means of genetic fuzzy systems (Cordón, Herrera, Hoffmann, & Magdalena, 2001), i.e., genetic algorithms (GAs) used to learn fuzzy rules. Two questions arise at this stage: *why fuzzy rules?* and *why GAs*?

The use of fuzzy rules (instead of other knowledge representations such as interval rules, decision trees, support vectors, neural networks and) is justified mainly by the kind of data set we are dealing with (see Section 3.1). In our case, each variable is composed of a set of parameters (items) that add uncertainty to the data, since each provides partial information to describe the variable. Moreover, we are able to transform with ease the available

expert knowledge into linguistic semantics. Finally, the obtained fuzzy models can be linguistically interpreted, an important issue in KDD.

Regarding the use of GAs to derive these fuzzy models instead of other well-known machine learning techniques, its application is justified by the following points. Firstly, since there are contradictory objectives to be optimized (such as accuracy and interpretability), we perform multiobjective optimization. It is one of the most promising issues and one of the main distinguishing marks of GAs as opposed to other techniques. Furthermore, we consider a flexible representation of fuzzy rules that can be properly developed by GAs.

In sum, the paper is organized as follows. Section 2 briefly describes the problem dealt with based on consumer behaviour models. Section 3 introduces the different KDD steps of the proposed methodology. Section 4 shows some experimental results obtained. Finally, in Section 5 we present the main concluding remarks of our research, as well as some final reflections.

2. Why should consumer behaviour modellers tend towards the use of KDD and artificial intelligence methodologies?

Marketing academics and practitioners have emphasized the need to know and explain consumer behaviour patterns in an increasingly efficient way. This is mainly due to firms focused on final markets being immersed in highly competitive systems in which their decision processes are required to be as accurate as possible. In this sense, models of consumer behaviour have proved throughout time to be a source of transcendental relevance for the development of the marketing science, as well as for the support of marketing managers' decision making (Van Bruggen & Wierenga, 2000).

However, recent contributions have strongly recommended the improvement and development of the marketing modelling discipline, with consumer behaviour modelling being part of it, with the aim of working with models that are more suitable and useful for academics as well as managers (Leeflang & Wittink, 2000). In this regard, there are several research issues whose improvement has been highlighted (Roberts, 2000; Steenkamp, 2000): the theoretical basis that supports and allows the proposal of the structure of models. This is needed in order to work with theoretical systems that are closer to the real system being modelled; and, secondly, the evolution of modelling estimation techniques and, in general, the methods used for the analysis of the relations among the constituent elements (variables) of the models. In this vein, in concordance with what we mentioned at the introductory part of this paper, we aim to contribute to the latter issue that we have just highlighted.

But why is it important to make advances in this issue? In our opinion, the answer is simple, though its solution implies thorough and diverse research efforts. Basically, it does not make any sense to evolve the theoretical framework of the consumer behaviour discipline, providing modern and suitable models that satisfactorily tackle the consumption problem under study, if such theoretic knowledge is not combined with proper methods of analysis capable of obtaining useful information from the market (i.e., the consumers' database).

The statistical techniques traditionally used to estimate current models of consumer behaviour do not seem to cover all the necessities to supposedly satisfy a method of analysis that aims to aid marketing decision making. This fact justifies why some researchers have highlighted the need to work with methods that are more accurate and oriented to the demand side in the near future (Gatignon, 2000). In other words, these methods must clearly fulfil the requirements of their users, usually marketing managers; i.e., being more complete, flexible and adapted to the strategic particularities of the competitive environment where the decision makers' firms operate. Likewise, as previously noted, inasmuch as the main problem currently faced by 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 (Van Bruggen & Wierenga, 2000). In this regard, considering the previous idea, some academics have predicted that in the mid term, in one or two decades, the MkMSS will tend to obtain benefits from integrating the modelling estimation techniques based on the classic econometric methods with the variety of expert systems based on artificial intelligence (Wedel et al., 2000).

Specifically, the methodology we propose, based on genetic fuzzy systems, can be associated with the incipient Fuzzy logic-based MkMSS (Li et al., 2000), and fulfils diverse requirements of the academics for future methods of analysis in marketing modelling; such as providing more flexible and interactive methods that offer a greater quantity of qualitative information than preceding estimation techniques traditionally used in this field (Gatignon, 2000).

3. A knowledge discovery method for consumer behaviour modelling with multiobjective genetic fuzzy systems

In this section, we introduce some of the main questions integrated in the KDD methodology that we propose be applied in the predictive analysis of consumer behaviour. In essence, as previously noted, we show and discuss the solutions we have given to the diversity of stages that basically constitute the KDD process, based on the marketing problem we face. However, those aspects related to the post-processing stage (interpretation), are treated in Section 4, where we show and analyze, with an illustrative orientation for the reader, the output of the method we use.

Specifically, the first three subsections dealt with questions related to the pre-processing stage. Thus, we tackle tasks like preparing the data or fixing the scheme we follow to represent the existing knowledge in the database. By

contrast, the rest of the subsections describe the machine learning method we have designed, based on GAs, in order to automatically extract the fuzzy models.

3.1. Data collection

First step is to collect the data related to the variables defining the theoretical model. In this sense, as has been traditionally done in marketing to analyze consumer behaviour, data are obtained by means of a questionnaire. Thus, at first, attention should be paid to how consumer behaviour modellers' face and develop the measurement process of the variables that such models contain. Its understanding is necessary in order to adequately approach the starting point of the KDD process, thus to give suitable and adapted solutions to the specific data we find in consumer behaviour modelling. In this regard, it can be said that the measuring streams for the constituent variables usually used in complex consumer models are classified in two differentiated groups, depending on if they defend that these constructs can or cannot be perfectly measured by means of observed variables (indicators) (Steenkamp & Baumgartner, 2000); i.e., the existence or not of a one-to-one correspondence between a construct and its measurement.

In the beginning, consumer behaviour modellers tended to make use of what was known as the *operational definition philosophy*. This is the simplest measurement approach, as it considered that there was a univocal correspondence between an element of the model – i.e., a construct/latent variable – and its measure. Consequently, on the basis of this measurement stream, there is no distinction between unobserved and observed variables in the measurement model, which is not considered very rigorous in the present day. On the other hand, a more convenient and reasonable position is that ulteriorly based on the *partial interpretation philosophy*, which distinguished between unobserved (constructs) and observed (indicators) variables. This latter approach of measurement, currently predominant in the marketing modelling discipline, poses the consideration of multiple indicators – imperfect when considered individually, though reliable when considered altogether – of the subjacent construct to obtain valid measures. Hence, we will take this measurement approach into account when dealing with how to process the data (see Section 3.2). Next, we show a simple example to illustrate this question.

For instance, we consider a simple measurement model depicted in Fig. 1, compounded by three construct or latent variables (depicted by circles), two exogenous and one endogenous, where: (1) *interaction speed*: the consumer's perception about the Internet's capacity in general, and, more particularly, of different web-sites, to give a response when required; (2) *invasion of privacy*: the consumer's opinion regarding the invasion of his/her intimacy by the various agents with which (s)he interacts in Internet applications; and (3) *attitude towards the Internet*: the consumer's overall attitude about this communications' medium. Since these latent variables cannot be directly measured, we measure indirectly them by means of observable variables (items), depicted by rectangles in the figure.

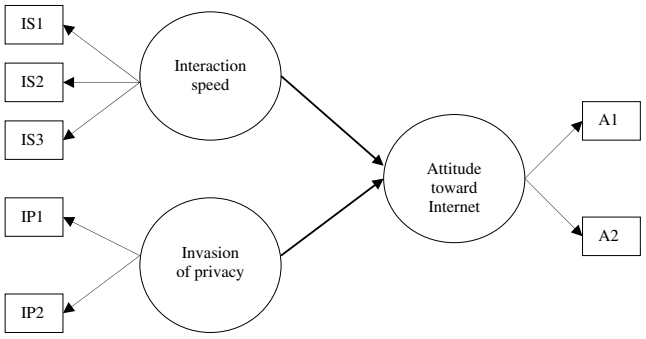

Fig. 1. Example of a simple consumer behaviour (measurement) model.

Likewise, with respect to the measurement scales, imagine, on the one hand, that the first and second constructs have been measured by means of several nine-point Likert scales ranging from 1: strongly disagree to 9: strongly agree. On the other hand, differential semantic scales with nine points have been used for the third. Specifically, in Table 1 we show a hypothetical example of the set of items – i.e., observed variables – that could have been used for measuring each one.

Finally, Table 2 shows an example of data available for this problem. There are just four cases, which is not realistic at all – i.e., think that a consumer database has usually hundreds or even thousands of individuals' responses gathered – though it is useful for our illustrative purpose. Note that the available data set consists of three variables, each composed of a set of values, thus providing an unusual kind of data.

Table 1
Example of a set of items related to the measurement model shown in Fig. 1

Interaction speed
IS1: Interaction with web pages is fast and stimulating
IS2: The internet is quick
IS3: Web pages that I usually visit download quickly enough
Invasion of privacy
IP1: When I surf the Internet, I feel my privacy has been invaded
IP2: Online firms do not respect the visitor's intimacy
Attitude toward the Internet
A1: Negative 1 2 3 4 5 6 7 8 9 Positive
A2: Unfavourable 1 2 3 4 5 6 7 8 9 Favourable

Table 2
Example of four consumers' responses about the items shown in Table 1

Cases	Interaction speed			Invasion of privacy		Attitude towards Internet	
	IS1	IS2	IS3	IP1	IP2	A1	A2
Consumer 1	2	3	2	6	7	2	2
Consumer 2	6	6	7	3	2	8	7
Consumer 3	8	8	9	2	3	9	9
Consumer 4	5	5	5	4	4	4	4

Furthermore, it is necessary to give some notes on the called "second-order constructs." These are a special case of latent variable that we may find in certain measurement models, though their use is not very common in consumer modelling (see Baumgartner & Homburg, 1996). Unlike the first-order construct, those latent variables with measures/indicators (items) directly related to them and just analyzed, these kinds of constructs do not have any direct associated indicator. Likewise, the second-order constructs are characterized by having several first-order constructs, usually dimensions of the former or combinations of several constructs, related to them. Basically, the second-order construct is inferred taking as a base the values of the first-order constructs associated with it. In other words, the first-order constructs act as their "indicators." Summarizing, as we may find measurement models, thus data sets, with these constructs, we deem it necessary to also reflect on and give a suitable solution to these cases in order to successfully apply the data mining process.

3.2. Data processing

Once the data have been collected, as explained in the previous section, it is necessary to adapt them to a scheme easily tractable by the machine learning algorithm. Therefore, our methodological 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. An intuitive approach could directly reduce the items of a specific variable to a single value (e.g., by arithmetic mean) (Casillas, Martínez-López, & Martínez, 2004). Another possibility would be to expand any multi-item example (the result of a questionnaire filled in by a consumer) to several single-item examples and subsequently reduce the data set size with some processes for instance selection.

Notwithstanding, 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 work with the original format without any pre-processing stage, and to give proper consideration to the existing uncertainty in the data (Martínez-López & Casillas, 2007): *the multi-item fuzzification*. Thus, a T-conorm operator (e.g., maximum), traditionally used in fuzzy logic to develop the union of fuzzy sets, is applied to aggregate the partial information given by each item during the inference process. Since it is not pre-processing data but a component of the machine learning design, the details of that treatment of the items are described in Section 3.4.2.

3.3. Representation and inclusion of expert knowledge

Several issues should be tackled at this step: the set of variables to be modelled, the transformation of marketing 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 her/his knowledge of the problem by means of a theoretic (measurement) model like that shown in Fig. 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, considering the measurement model shown in Fig. 1, the basic structure of a fuzzy rule presents the following form:

IF *Interaction Speed* is A_1 and *Invasion of Privacy* is A_2
THEN *Attitude towards Internet* 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 the expert builds the questionnaire, in order to collect the data necessary, she/he must fix the kind of scale and precision (number of points) used to measure each variable. Thus, considering this prior 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 scales (i.e., Likert differential semantic or rating scale), which is one of the most commonly used in marketing, so giving an *ad hoc* solution to them.

Specifically, we suggest transforming these scales into Ruspini's strong fuzzy semantics with uniform density of the fuzzy membership functions in order 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:

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

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

Fig. 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*.

At this stage, one could think about using some mechanism to automatically generate fuzzy partitions from data

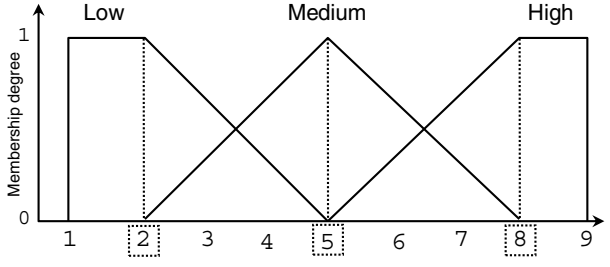

Fig. 2. Transformation of a nine-point rating scale into a fuzzy semantic.

(Guillaume & Charnomordic, 2004), or design a tuning method to adapt uniformly initialized fuzzy semantics (Casillas, Cordón, del Jesus, & Herrera, 2005), or both of them. However, considering the problem we face, the expert is not interested in generating fuzzy semantics that accurately cover the data. On the contrary, it must be taken into account the fact that the truest way to interpret the semantic considered by each consumer who filled the questionnaire is the uniform one. Consequently, if we apply any automatic process to generate/tune fuzzy membership functions, we would be adapting to the context – i.e., the answers of the consumer – but not to the meaning of the variables. Therefore, in this problem the KDD process is focused on the relationship among the variables (fuzzy rule surface structures).

3.4. Data mining

Once the linguistic variables that properly represent the information requested from the consumer are fixed, a machine learning process must be used to automatically extract the existing knowledge in the data. This task is, without doubt, the core issue from the KDD point of view. As mentioned in the introductory part, this paper is interested in predictive induction. Of course, since we are performing knowledge discovery, the model obtained should not only be accurate enough but also be easily legible in order to be able to describe the real system linguistically. As is known, accuracy and interpretability are two contradictory properties. Then, to properly address that, we choose multiobjective genetic fuzzy systems thanks to their good behaviour in dealing with multiple, contradictory objectives. The following sections describe the main components of the proposed method.

3.4.1. Fuzzy rule structure

In data mining, it is crucial to use a learning process with a high degree of interpretability. Therefore, we opt for a compact description of the antecedent by expressing it in normal conjunctive form. This kind of fuzzy rule structure is commonly known as a DNF-type fuzzy rule (González & Pérez, 1998). This kind of fuzzy rule structure has the following form:

$$R: \text{IF } X_1 \text{ is } \widetilde{A}_1 \text{ and} \ldots \text{and } X_n \text{ is } \widetilde{A}_n \text{ THEN } Y_1 \text{ is } B_1 \text{ and} \ldots \text{and } Y_m \text{ is } B_m$$

where each input variable X_i, $i \in \{1,\ldots,n\}$, taking as a value a set of linguistic terms $\widetilde{A}_i = \{A_{i1} \text{ or} \ldots \text{or } A_{ini}\}$, whose members are joined by a disjunctive (T-conorm) operator, whilst the output variable remains a usual linguistic variable with a single associated label. We use the *bounded sum* as T-conorm in this paper:

$$\mu_{\widetilde{A}_i}(x) = \min\left\{1, \sum_{k=1}^{n_i} \mu_{A_{ik}}(x)\right\}$$

The structure is a natural support to allow the absence of some input variables in each rule (simply making $\widetilde{A}_i$ to be the whole set of linguistic terms available).

3.4.2. Multi-item fuzzification

In order to properly consider the set of items available for each input/output variable (first-order construct) as discussed in Section 3.2, we propose an extension of the membership degree computation, the so-called multi-item fuzzification. The process is based on a union of the partial information provided by each item. Given X_i and Y_j measured by the vectors of items $\vec{x}_i = (x_1^{(i)},\ldots,x_{h_i}^{(i)},\ldots,x_{p_i}^{(i)})$ and $\vec{y}_j = (y_1^{(j)},\ldots,y_{t_j}^{(j)},\ldots,y_{q_j}^{(j)})$, respectively, the fuzzy propositions "X_i is $\widetilde{A}_i$" and "Y_j is B_j" are, respectively, interpreted as follows:

$$\mu_{\widetilde{A}_i}(\vec{x}_i) = \max_{h_i=1}^{p_i} \mu_{\widetilde{A}_i}(x_{h_i}^i)$$

$$\mu_{B_j}(\vec{y}_j) = \max_{t_j=1}^{q_j} \mu_{B_j}(y_{t_j}^{(j)})$$

Therefore, the T-conorm of *maximum* is considered to interpret the disjunction of items.

Furthermore, as mentioned in Section 3.1, second-order constructs/variables are characterized by not being associated directly to items, on the contrary they are inferred by taking its associated first-order constructs as reference. Since second-order constructs represent the intersection of several first-order constructs, we can consider a T-norm as fuzzy conjunction to gather the information given by each first-order construct. Therefore, given the second-order input variable $X_i = \{X_{i1},\ldots,X_{ik_i},\ldots,X_{is_i}\}$ – with s_i being the number of independent first-order variables – the fuzzy proposition "X_i is $\widetilde{A}_i$" is interpreted as follows:

$$\mu_{\widetilde{A}_i}(\vec{x}_i) = \min_{k_i=1}^{s_i} \mu_{\widetilde{A}_i}(\vec{x}_{ik_i})$$

Therefore, the T-norm of *minimum* is considered to interpret the conjunction of first-order constructs in this paper.

3.4.3. 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 [small], M [medium], and L [large]) 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.

3.4.4. Objective functions

We consider three objective functions to assess the quality of the generated fuzzy systems, one based on the approximation error to optimize the accuracy and two others based on the linguistic complexity to optimize the interpretability.

- *Approximation error* (F_1): This measure refers to the capability of the generated fuzzy model to faithfully represent the real-world system (expressed by the data set). The closer the model to the system, the lower its error. Since the output variable is a composition of several items (see Table 2 and Section 3.2), we have adapted the root mean square error (RMSE) computation to consider that. As mentioned before, the aggregation of items is made by the union. So, let us suppose that the output variable is composed of two items and the prediction had a success degree of S_1 for the first item and S_2 for the second. The total success degree would be $S_1 \vee S_2$. From De Morgan's laws, $S_1 \vee S_2 = \overline{\overline{S_1} \wedge \overline{S_2}}$, i.e., $\overline{E_1 \wedge E_2}$ – with E_1 and E_2 being the errors (complement of the success degrees) done over the corresponding items. If the objective of the algorithm is to maximize $S_1 \vee S_2$, the complement will be to minimize $E_1 \wedge E_2$. Therefore, the objective function (for minimization) in MISO (multiple-input, single-output) system is as follows:

$$F_1(\mathrm{FS}) = \sqrt{\frac{1}{N}\sum_{e=1}^{N} \min_{t=1}^{q} (\mathrm{FS}(x^{(e)}) - y_t^{(e)})}$$

  with FS being the evaluated fuzzy system, $(\mathbf{x}^{(e)}; \vec{y}^{(e)})$ being the eth example, $\mathbf{x}^{(e)} = (\vec{x}_1^{(e)}, \ldots, \vec{x}_n^{(e)})$ the input item vectors, and $\vec{y}^{(e)} = (y_1^{(e)}, \ldots, y_q^{(e)})$ the output item vector. Notice that $\mathrm{FS}(\mathbf{x}^{(e)})$ performs the fuzzy inference using the multi-item fuzzification described in Section 3.4.2.
- *Number of DNF-type fuzzy rules* (F_2): This second objective assesses the size of the generated fuzzy rule set. It is clear that the higher number of rules, the greater the complexity and the worse the interpretability. The objective consists of minimizing the number of fuzzy rules contained in the fuzzy system:

  $$F_2(\mathrm{FS}) = |\mathrm{FS}|$$

- *Number of equivalent Mamdani fuzzy rules* (F_3): The objective F_2 does not completely assess the linguistic complexity of the fuzzy system since the internal structure of each DNF-type fuzzy rule is not considered. Therefore, in order to seek out fuzzy rules as general and simple as possible, we include a third objective that measures the mean number of equivalent Mamdani fuzzy rules for each DNF-type fuzzy rule as follows:

$$F_3(\mathrm{FS}) = \frac{1}{|\mathrm{FS}|}\sum_{R_r \in \mathrm{FS}} \prod_{i=1}^{n} l_{ri}$$

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 in a rule (i.e., "don't care"). The objective is to maximize this third objective in order to generate more general fuzzy rules.

3.4.5. Evolutionary scheme

A generational approach with the multiobjective NSGA-II replacement strategy (Deb, Pratap, Agarwal, & Meyarevian, 2002) is considered. Crowding distance in the objective function space is used; in each front, this measure is normalized from the minimum and maximum values for each objective in that front. Binary tournament selection based on the non-domination rank (or the crowding distance when both solutions belong to the same front) is applied.

3.4.6. Genetic operators

The *crossover* operator randomly chooses a cross point between two fuzzy rules at each chromosome and exchanges their right string. 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 cannot be applied if only one 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 that appeared after mutation are subsumed by the most general ones and redundant rules are removed.

3.4.7. Inference mechanism

When using DNF-type fuzzy rules, special care must be taken with the inference engine. Indeed, for a proper behaviour of the algorithm, it is mandatory to ensure that they are also numerically equivalent, given two linguistically equivalent rule bases. In order to do so, we consider the FATI (first aggregate, then inference) approach, the Max–Min scheme (i.e., T-conorm of maximum as aggregation and T-norm of minimum as implication operator), T-norm of minimum as conjunction, and centre-of-gravity as defuzzification.

4. Experimentation with a consumer behaviour modelling application

In order for the reader to follow a logical and understandable sequence of contents, we have deemed it necessary to structure this section in three parts. First, we make some minimal commentaries about the model and database used to empirically show how KDD methodology for consumers' predictive modelling works. In this respect, we are aware that this is a secondary question, inasmuch as what is relevant is not the theoretical basis of model used, but the potential and performance of our methodology. Nevertheless, a brief description of the aim and scope of the model used for the experimentation will help us to see better the relations among variables we consider in this application. Second, as this is a new KDD method, we also comment on some of the main steps that should be followed in order to make a correct interpretation of its output. In other words, we synthetically present the main questions to be tackled in what may be considered an analysis protocol; this is very helpful in addressing the post-processing (interpretation) KDD stage correctly. Finally, once the previous issues have been presented, we then dedicate the last section to show and discuss a variety of results obtained after applying our predictive modelling method to the data.

4.1. Application model and data: previous comments

The consumer behaviour model we have used for the experimentation of our KDD methodology is based on a causal model already proposed by Novak, Hoffman, and Yung (2000), whose central element is consumer's *flow state* when surfing the Web. As the authors allow the use of their database for academic purposes, we have opted to experiment our methodology on a consumer model already validated and widely known among academics in the marketing field. This is a plausible and orthodox alternative, as we can see by analyzing other research previously developed (e.g., Beynon, Curry, & Morgan, 2001; Fish, Johnson, Dorsey, & Blodgett, 2004; Hurley, Moutinho, & Stephens, 1995; Levy & Yoon, 1995; Rhim & Cooper, 2005; Yi-Hui, 2007), when proposing and testing new KDD applications in marketing.

In order to briefly introduce this concept, so the reader better understands the variable we want to explain in this empirical application of our methodology, we now synthetically present some ideas about it. Flow has been recently imported from motivational psychology and successfully adapted to explain consumer behaviour phenomena on the Web (Hoffman & Novak, 1996; Korzaan, 2003; Luna, Peracchio, & De Juan, 2002; Novak et al., 2000; Novak, Hoffman, & Duhachek, 2003). In general terms, flow state is defined as "the process of optimal experience" or the mental state that individuals sometimes experience when they are deeply immersed in certain events, objects or activities (Csikszentmihalyi, 1975, 1977). This concept has been adapted to the Web environment. In this context, flow state is achieved when the consumer is so deeply involved in the process of navigation on the Web that "nothing else seems to matter" (Hoffman & Novak, 1996, p. 57).

The model we consider for the experimentation has 12 elements (constructs) causally related. In this regard, we have transformed this causal model into a hierarchical fuzzy system (see Fig. 3), formed by six fuzzy rule-based systems (FRBS) interconnected; one FRBS for each of the consequents (endogenous elements) of the model of reference. In other words, the design of the system has been done in such a way that we are able to predict the consumer's behaviour with respect to the set of dependent variables of the model, considering their multiple interrelations.

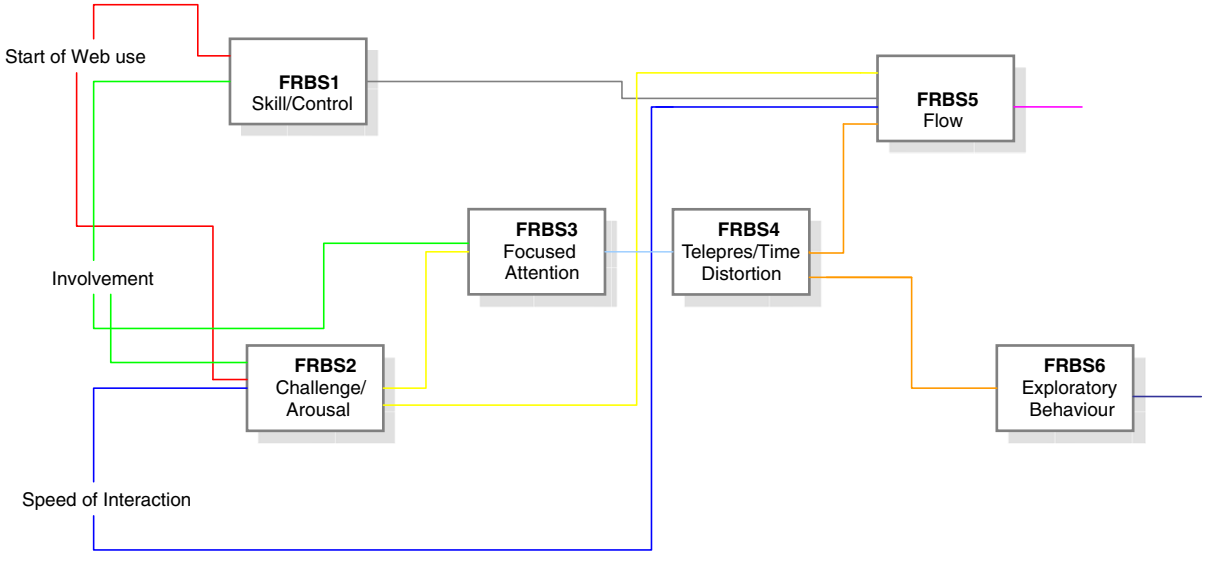

Fig. 3. Hierarchical fuzzy system associated with the marketing model used in the application of the methodology.

Specifically, due to space constraints, in this paper we focus on showing consumers' behaviour predictions for two consequents of the set of constituent endogenous elements of the system, with the aim of showing the performance of our KDD method. Next, we introduce some minimal theoretic notes about them, though we suggest consulting Novak et al. (2000) if a deeper understanding of the theoretical basis of the systems than used in our application is required. These are the following:

- Predicting *focused attention (FRBS3)*. This system is formed by two antecedents: *involvement* with the Web and the second-order construct *challenge/arousal*. The former represents how important the Web is for the user, while the latter gathers how challenging and stimulating surfing the Web is. Both predictors were theoretical hypothesized to exert a positive influence on the level of attention shown by a user to his/her process of navigation.
- Predicting *flow state (FRBS5)*. This system considers the four primary antecedents of the consumer's Flow State (consequent). Specifically, based on theoretical relations of the model of reference.
  - *Speed of Interaction* refers to the user's perception of how quick the process of interaction is when using the Web.
  - *Skill/Control* gathers the consumer's opinion regarding his/her own capacity to develop successful navigating processes on the Web.
  - *Telepresence/Time Distortion* is also a compound construct that refers to the consumer's perception about the predominance of the virtual computer (Web) environment over the physical environment where the consumer is placed when surfing the Web, as well as to the loss of the consumer's self-consciousness regarding the notion of time when developing such a process of navigation.
  - *Challenge/Arousal*, already commented on in the previous paragraph. These four elements have been hypothesized to exert a positive relation on the consumer's Flow State.

Most parts of the construct, except one which was measured by means of an ordinal scale, were gathered by multi-item Likert scales with nine points; i.e., metric scales. Specifically, the fuzzy semantic we have applied to all the elements of the systems introduced above is shown in Fig. 2.

4.2. Steps to follow when post-processing the results: protocol of analysis

In KDD, the post-processing of the results generated in the data mining stage is also very important to achieve a successful application of the KDD process; hence, to obtain valuable information about the problem to be solved. Moreover, this question of the interpretation of the results should be especially considered in brand new KDD methodologies, as is the case in this paper. In other words, the steps followed to provide meaning to the output obtained via the application of the machine learning stage must not respond to an *ad hoc* improvisation of the expert, which will likely vary in each situation where the KDD method is applied. On the contrary, it is convenient that such structural steps are the result of a systematic procedure specifically designed on the basis of the kind of knowledge that the method aims to provide.

In this section, we propose a general procedure to be followed when analyzing the results coming from the machine learning stage of our methodology. Specifically, inasmuch as the method we present here has been designed for predictive marketing modelling, the application of this protocol will facilitate knowledge extraction for two core questions: first, the sense of the causal relations among the elements of the model; and second and most useful, specially from a professional perspective, the behaviour/evolution of the predicted variable, as its predictors vary their values. The latter question is plausible thanks to the capabilities of this KDD method, which allows predictions to be made regarding the variables of interest by simulating diversity of scenarios for the predictor variables formed in every FRBS.

Next, we briefly present the protocol of analysis we propose:

Step 1. Graphical representation and analysis of the Pareto front. This graphical illustration allows the representation of the set of solutions (i.e., hierarchical fuzzy systems) obtained, after the application of the algorithm, by means of a Pareto front; this kind of representation is highly reasonable and coherent with the multiobjective optimization algorithm we apply. Specifically, the aim of using such a graph is twofold: first, it allows visualization of both the accuracy and the degree of difficulty in interpreting each system generated; second, it permits the selection of the hierarchical fuzzy system that is most suitable, based on the accuracy/interpretation trade-off of every system generated by the algorithm. In this sense, as what should be more valued is the predictive accuracy of the system, we generally recommend selecting the one that shows the minimum error. However, the marketing expert may find a particular system in which, though its error is slightly higher than the optimum in terms of accuracy, the interpretability is much more interesting as it is less complex (it is compounded by a significant inferior number of general rules). In such cases, the expert should analyze which is the best system to maintain, jointly considering accuracy and legibility, based on his/her predictive objectives.

Step 2. Analysis of the transference function. Basically, this function is the graphic surface that relates the set of elements (variable) integrating certain FRBS. Specifically, it offers a global vision of the behaviour of the variable to be predicted in function of the hypothetical values that the predictor(s) may take. The main aim that justifies and makes the use of transference functions convenient lies in the essence of how the hierarchical fuzzy systems are

extracted. In simple terms, when using predictive induction, each constituent rule of a FRBS is part of a global solution, so it would not be orthodox to analyse each rule separately. On the contrary, this would be normal if using descriptive induction, but, as commented in Section 1.2, it is not a valid part of the research aims in this paper. However, such a rule is necessary for the cooperative set of rules integrating certain FRBS, which may globally be the best solution. In other words, when interpreting the relations among variables, the expert must support his/her conclusions primarily on the visual analysis of the graphic transference function. Likewise, the expert may draw conclusions from certain rules of the FRBS as a secondary route, but without losing the global vision of the system. In this regard, using only the semantic representation of the set of rules as a reference is difficult for the "human eye." This is why the graphical transference function is so useful, especially as the FRBS are more complex.

With respect to this, we deem it necessary to make some reflections on the graphical representation of the transference functions. First, if the FRBS has just one predictor, all the information of the system can be visualized in just one graph; this is the simplest case. In this scenario, the transference function is graphically represented on the plane. Likewise, when the FRBS has two predictive variables, we have followed the underlying philosophy of the three-dimensional graphical representations by matching a chromatic scale to the range of variation of the original marketing scales.

But, and this may be the more challenging question, what to do when the FRBS has more than two variables–predictors? The solution is more complex for this scenario, as the visualization of the whole FRBS is not possible with just one graph. We have solved this question by means of what we call "chromatic transition maps."

In essence, as the number of predictors is higher than two (i.e., the maximum number that would allow the representation of the whole FRBS), it is convenient that the expert, based on aprioristic information, selects the two predictors which are more relevant; this will facilitate the visual analysis. Then, the values of the rest of predictors should be fixed. This is due to a simple reason: inasmuch as the main objective is to find tendencies in the relations among predictors and the variable to be predicted, the values of the rest of the predictors should be iteratively modified with the aim of analyzing the surfaces' evolution of the graphical transference function. This question will be illustrated in the next section.

Once the questions introduced in the paragraph above have been treated, the graphs of the transference functions can be generated. Hence, every value we fix for each of the predictors, apart from the two initially selected, will have a transference that is graphically related to it. Therefore, we will work with as many graphs as points we fix for such predictive variables. In our methodology, we have called this set of graphs "chromatic transition maps." In this sense, unlike the three-dimensional graphs associated with FRBS with two predictors whose graphical representation is more versatile (i.e., they can be rotated), these maps are characterized by their representation of each graphical transference function with a vertical, up/down perspective. In this regard, we can easily see the relations among variables by observing the chromatic evolution of the surfaces on the graph.

Step 3. Presentation of the FRBS. Lastly, the information offered by the FRBS, in a symbolic structure, may be used after the visual analysis of the transference functions and the chromatic transition maps. In any case, at this stage of the protocol, what justifies examination of the constituent rules of the FRBS is either confirmation of certain tendencies identified in the previous stage or, more specifically, clarification of the analysis of certain relations among the elements of the system.

4.3. Experimental results and knowledge interpretation

Training data are composed of 1154 examples (consumers' responses). We have run the algorithm 10 times, obtaining the following values for the parameters: 300 generations, size of the population 100, crossover probability 0.7, and the probability of mutation per chromosome 0.1.

4.3.1. Predicting focus attention

Our algorithm has generated seven alternative FRBS, with different degrees of accuracy and interpretability; see the Pareto front related to this system in Fig. 4. As recommended, it is convenient, taking into account our predictive purposes, to work with the most accurate system. In this case, we have chosen the FRBS with three DNF and eight Mamdami fuzzy rules.

Now, once we have decided which FRBS to use in order to better explain the relations between the predictors (i.e., importance and challenge/arousal) and the consumer's focus attention when surfing the Web, the next step is to generate and analyze its transference function. The system

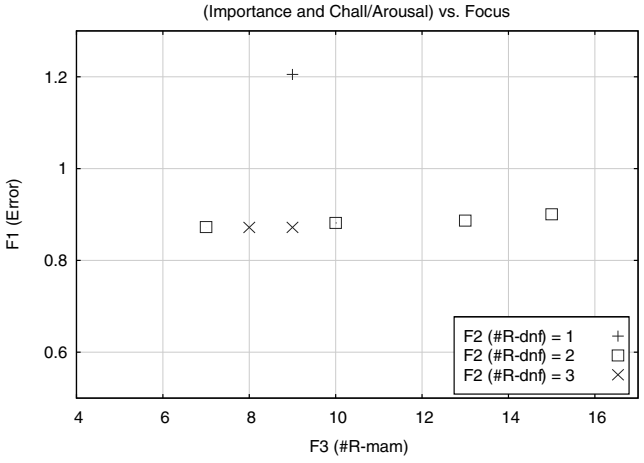

Fig. 4. Pareto front of the alternative FRBS to predict focus attention.

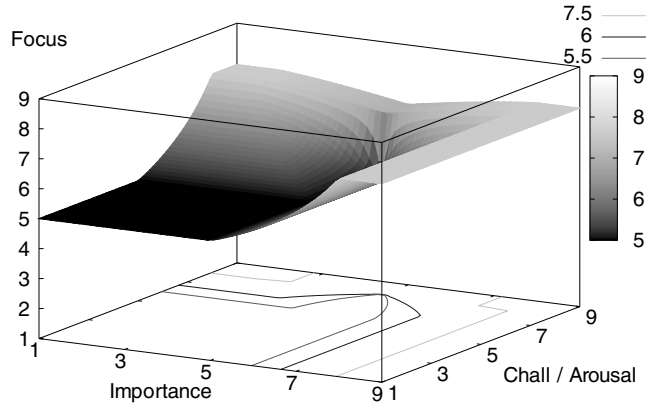

Fig. 5. Transference function associated with the FRBS selected to predict focus attention.

to be represented is compounded by three variables, so we obtain a three-dimensional function, as shows in Fig. 5.

The information offered by the transference function is very interesting and clear. First, both predictors have a positive influence on the consumer's focus attention. However, such influence is neither linear nor constant along the set of values that may take the predictors. This question is significant as the statistical method usually used to estimate these kinds of models, hence to analyze the relations between variables, are usually constrained by a linear parameter; for instance, the model we are using to apply our methodological proposal was estimated via structural equation modelling. In this sense, the predictive modelling we have presented in this paper allows us to reach much higher levels of qualitative information about how the system (i.e., variables relationships) behaves. Such information is very helpful for supporting the market decisions managers usually take.

The graphical results provided by the simulation we have developed are illustrative in this respect. It is expected that a consumer will present moderate levels of focus attention when navigating, even if both his/her involvement with the Web and perception about how challenging surfing the Web are low. In other words, we could generally expect that consumers, when developing an online process of navigation, should be, as a minimum, moderately focused. This is not unreasonable, as surfing the Web is characterized by being an interactive process, where the consumer must have a minimum level of concentration in order to follow a coherent sequence of movements (i.e., click streams).

However, the transference function clearly suggests that the consumer's focus attention when surfing the Web will increase from "moderate" to "high" as either of their two predictors, jointly or separately, take values "moderate" or higher. In other words, both predictors will not produce significant variations on the consumer's focus attention when they take values from "low" to "moderate," though they will boost his/her level of concentration as their levels go beyond. Specifically, and this is an interesting point, though the highest gradient for Focus Attention is when Importance and Challenge/Arousal are simultaneously high, each predictor shows a similar pattern of influence in Focus Attention as it goes from "low" to "high," even if the other predictor takes low levels. That is to say, it is expected that high levels of consumer involvement with the Web will also produce good states of consumer focus attention, even when the Challenge/Arousal perception is low, and vice versa. Thus, it seems that there is not much interaction between either predictor when determining the consumer's level of focus attention.

Finally, in Table 3 we show the three constituent DNF-type fuzzy rules of the FRBS selected. In non-complex systems, as is the case of predicting focus attention, it is usually easy to see any correspondence between the visual analysis of the transference function and the fuzzy rules integrated in the system.

4.3.2. *Predicting flow state*

In this case, we follow an equal structure of analysis. The particularity, however, and this is why we have also decided to illustrate it in this paper, is that we work with a more complex FRBS, in which "transition chromatic maps" should be used. This can be considered a visual modelling process that represents the extracted knowledge in a more understandable way, thus helping in the post-processing, interpretation stage of KDD.

Table 3
Set of rules associated with the transference function shown in Fig. 5 – mdm stands for medium

Importance			Challenge/Arousal			Focus attention		
Low	Mdm	High	Low	Mdm	High	Low	Mdm	High
×	×		×	×			×	
×					×			×
		×						×

F1 (RMSEtra): 0.871808, RMSEtst: 0.747576, F2 (number of DNF-type fuzzy rules): 3, F3 (number of equivalent Mamdani fuzzy rules): 8.

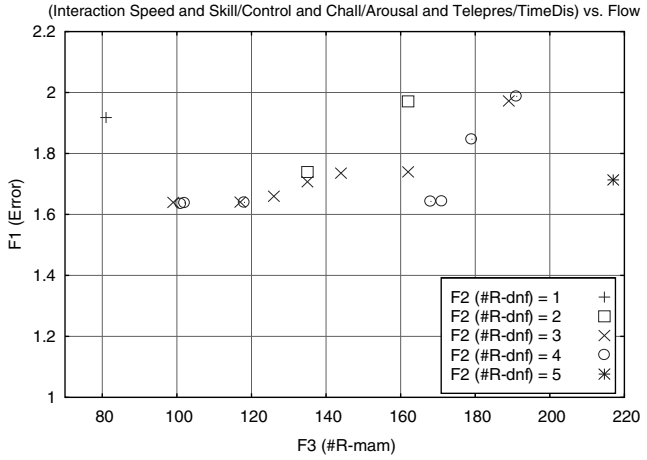

Fig. 6. Pareto front of the alternative FRBS to predict flow.

It is reasonable that, as the system increases its complexity, the algorithm works with a higher number of alternative FRBS during the machine learning stage. The case we analyze now is characterized by four variables to predict the consumer's Flow State. In Fig. 6, we show the Pareto front with the plots/FRBS for the three objectives under consideration. The FRBS finally selected is the one with four DNF and 101 Mamdami fuzzy rules. This is the most accurate from the whole set of alternatives generated.

As commented in Section 4.2 (Step 2), when we work with more than two predictors, as in the case of this FRBS, we have to make use of chromatic transition maps. Although this is a more complex graphical representation of the transference functions associated with an FRBS, it is a useful solution to achieve a global vision of the system's behaviour. This FRBS has five variables, four predictors vs. flow state, as shown in Fig. 7. If the reader recalls, the organization of the predictors on the map is not random. On the contrary, as we have at our disposal *a priori* information – i.e., parameters of the model estimated by structural equation modelling (SEM), in Novak et al.'s (2000) paper – we have selected Telepresence/Time Distortion (hereafter TP/TD) and Challenge/Arousal in every graph representing the transference functions associated with a particular scenario of values fixed for the other two predictors, *a priori* less influential, Interaction Speed and Challenge/Arousal.

Likewise, the position of the range of values (i.e., from 1 to 9) for every predictor in the transference functions' axes has been set in such a way that facilitates the visual analysis of Flow. Specifically, this configuration of the axes allows a diagonal view up-to-down, left-to-right, to relate to the consumer's Flow States associated with increasingly higher values for the predictor' variables.

Though the visual analysis of a chromatic transition map may present certain difficulties, at first, for an

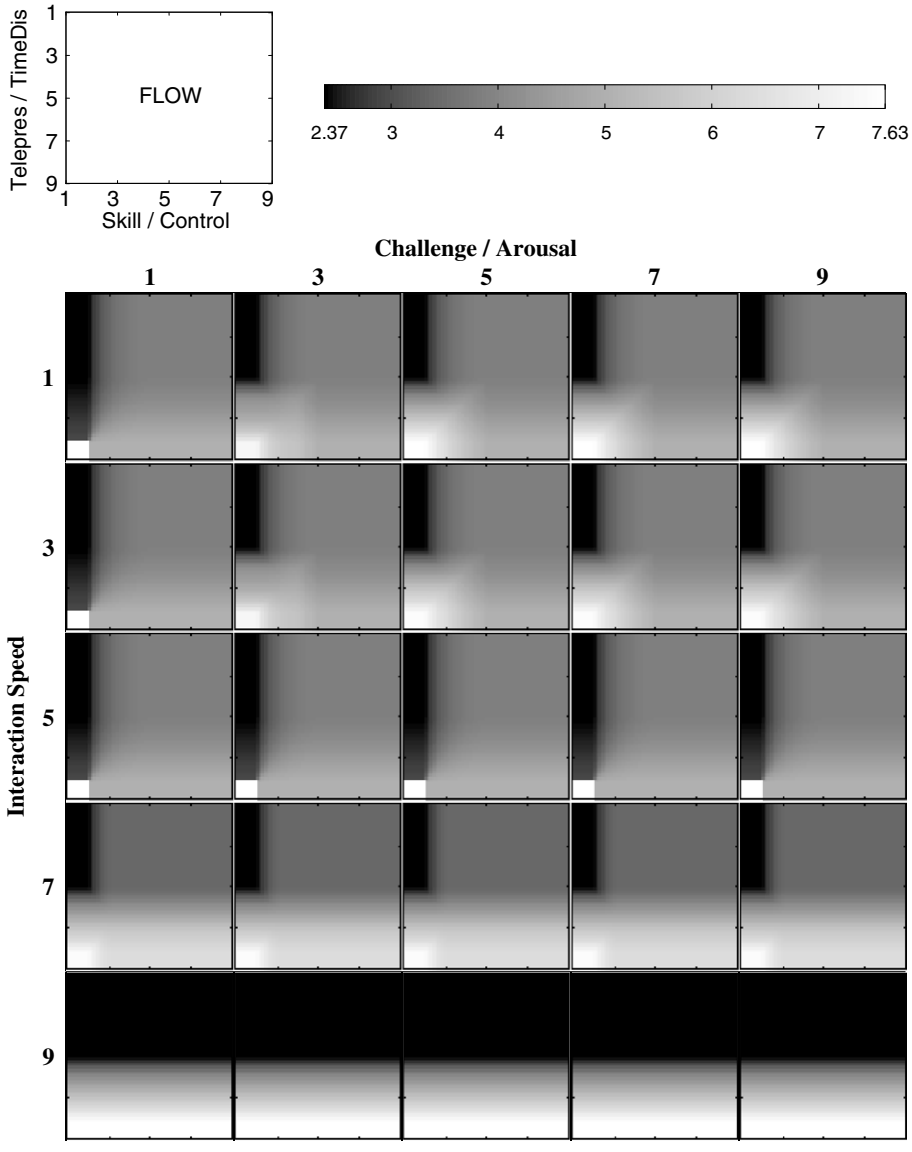

Fig. 7. Chromatic transition map generated to predict the consumer's Flow State.

Table 4
Fuzzy rule set associated with the chromatic transition map (transference function) shown in Fig. 7 – mdm stands for medium

Interaction speed			Skill/Control			Chall/Arousal			Telepress/TimeDistortion			Flow		
Low	Mdm	High	Low	Mdm	High	Low	Mdm	High	Low	Mdm	High	Low	Mdm	High
									×	×		×		
×	×			×	×								×	
×			×				×	×			×			×
		×									×			×

F_1 ($RMSE_{tra}$): 1.635246, $RMSE_{tst}$: 1.650746, F2 (number of DNF-type fuzzy rules): 4, F3 (number of equivalent Mamdani fuzzy rules): 101.

"untrained eye," its utility is evident as the marketing expert achieves appropriate levels of familiarity with this predictive modelling tool. Next, we present some notes to illustrate the kind of information to be obtained from this particular way of representing the transference functions.

In theory, the four predictors should be positively related to the consumer's Flow State. In this sense, it is demonstrated by an empirical estimation of such relations by SEM. However, the consumers' database would be underused if the expert did not try to go beyond that. We insist again on the valuable qualitative information that methods like the ones presented in this paper may provide the marketing manager. Let us see what would happen to the FRBS we have tried here, in order to predict Flow.

Skill/Control helps to explain the consumer's Flow State transitions from low to moderate levels as it increases. However, this predictor saturates its influence when taking moderate or higher levels. In any case, there are some exceptions to this general idea that are worth commenting on.

TP/TD varies its degree of influence depending mainly on the consumer's perception about the Speed of Interaction with the Web. In fact, Speed of Interaction seems highly significant in discriminating between two clear scenarios of influence for the other three predictors.

Taking the consumer's opinion about Speed of Interaction to be poor, it is expected that TP/TD will not be very influential in determining high Flow States, except when this predictor is high. In this case, TP/TD considerably amplifies the consumer's Flow levels, with more intensity as the consumer perceives that surfing the Web is more challenging. Thus, there seems to be a clear positive interaction between TP/TD and Challenge/Arousal.

But, probably, one of the most interesting conclusions we can draw from this scenario is about the variation in the sign of the influence of Skill/Control in the consumer's Flow State. Paradoxically, as theoretically expected, Skill/Control is positively related, as previously commented, and such a relation is reversed when the consumer experiences high TP/TD. This finding is more evident as Challenge/Arousal increases. Therefore, in these circumstances, it is expected that the consumer's Flow State evolution will be inhibited by better capabilities to surf the Web. The explanation for this apparent paradox is logical. In general, an individual should be more concentrated on what he/she is doing, in this case an online navigational process, as he/she becomes better qualified/prepared to do it, and vice versa. However, considering the electronic context we are analyzing, when the individual is really involved in the experience of the virtual world, thus presenting high levels of TP/TD, it is expected that both will outshine and offset his/her poor technical capabilities in properly surfing the Web, in order to have high Flow States. In other words, based on our results, TP/TD is probably the most important factor, when it is high, to predict high levels of Flow in consumers. Notwithstanding that, it is expected that this special effect on the consumer's mind state caused by being really hooked on the virtual experience will be reduced as he/she acquires greater control, hence more consciousness, over what he/she is doing.

On the other hand, when the consumer's perception about Speed of Interaction is medium and, especially, high, we identify another scenario of influence. In this case, Challenge/Arousal does not have any predictive capacity to explain Flow State. This is easy to conclude when obtaining the same surface graph for the transference function in each of the five points fixed for this variable. Also, Skill/Control shows a poor influence over Flow, or non-existent when Speed of Interaction is maximum. Without doubt, the most determinant predictor of Flow State is TP/TD. In this case, one can also observe how high levels of Speed of Interaction and TP/TD predict high Flow states, regardless of the value taken by the other two predictors.

Finally, in Table 4 we show the four DNF rules belonging to the FRBS selected. This system is complex, which makes it less convenient to analyze each rule separately; the reader should remember the spirit that determined the generation of the set of rules. However, taking this question into account, some of the rules are very interesting for corroboration of some of the main conclusions we have commented on after the visual analysis of their chromatic transition maps. For instance: the role of TP/TD in predicting high Flow States in consumers, or the non-influence of Challenge/Arousal and Skill/Control when Speed of Interaction is high.

5. Concluding remarks

The paper has introduced a novel problem in marketing where KDD can help to generate easily understandable models for predictive induction. As far as we know, this

is the first time that KDD has been applied to estimate structural models for consumer behaviour, which is usually done by traditional statistical tools.

The proposed methodology develops three different stages of KDD: data collection, data mining, and knowledge interpretation. Data are collected from questionnaires based on a theoretically defined structural model filled in by consumers. The proposed data mining approach is based on the use of genetic algorithms to learn fuzzy rules. The problem provides a specific kind of uncertain data set that justifies the use of fuzzy logic. We perform multiobjective optimization (according to several quality criteria) to obtain diverse fuzzy models with different balances between accuracy and legibility. These alternative solutions can be analyzed by an expert from plots that collect the considered quality criteria. Finally, the solutions selected are interpreted by means of visual modelling that shows the system behaviour in a graphical and compact way, thus helping the expert to take decisions about the market analyzed according to the consumer's opinions.

Our KDD methodology has been appropriately applied to a real-world problem that analyzes consumer behaviour in interactive computer-mediated environments. As further work, we intent to extend our KDD approach to other areas in social science that use similar kinds of data to analyze human behaviour.