

Marketing Intelligent Systems for consumer behaviour modelling by a descriptive induction approach based on Genetic Fuzzy Systems

Francisco J. Martínez-López^{a,*}, Jorge Casillas^{b,1}

^a Department of Marketing, Business Faculty, University of Granada, Granada, E-18071, Spain

^b Department of Computer Science and Artificial Intelligence, Computer and Telecommunication Engineering School, University of Granada, Granada, E-18071, Spain

ARTICLE INFO

Article history:

Received 2 March 2007

Received in revised form 26 December 2007

Accepted 12 February 2008

Available online 14 April 2008

Keywords:

Marketing modelling

Management support

Analytical method

Knowledge discovery

Genetic Fuzzy Systems

Methodology

ABSTRACT

In its introduction this paper discusses why marketing professionals do not make satisfactory use of the marketing models posed by academics in their studies. The main body of this research is characterised by the proposal of a brand new and complete methodology for knowledge discovery in databases (KDD), to be applied in marketing causal modelling and with utilities to be used as a marketing management decision support tool. Such methodology is based on Genetic Fuzzy Systems, a specific hybridization of artificial intelligence methods, highly suited to the research problem we face. The use of KDD methodologies based on intelligent systems like this can be considered as an avant-garde evolution, exponent nowadays of the so-called knowledge-based Marketing Management Support Systems; we name them as Marketing Intelligent Systems. The most important questions to the KDD process—i.e. pre-processing; machine learning and post-processing—are discussed in depth and solved. After its theoretical presentation, we empirically experiment with it, using a consumer behaviour model of reference. In this part of the paper, we try to offer an overall perspective of how it works. The valuation of its performance and utility is very positive.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Firms operate in markets that are increasingly “turbulent” and “volatile.” How to deal with this turbulence and survive in these hypercompetitive conditions has become a strategic question (Agarwal, Shankar, & Tiwari, 2007; Christopher, 2000). Consequently, the idea of the achievement and support of a sustainable competitive advantage gave rise, in the nineties, to another focused on its continuous development (D’Aveni, 1994), which is more realistic these days. One of the main implications of this reformed strategic approach is a search for new suitable market opportunities. Of course, such opportunities need to be correctly identified and addressed by firms. This premise justifies the transcendental relevance recently given to the creation and management of knowledge about markets (Drejer, 2004). In this vein, the marketing function of companies and, most especially, their Marketing Management Support Systems (MkMSS) plays a notable role in this task, as they must contribute to the reduction of the uncertainty related to the firms’ markets of reference. As we know, this question does not only imply having access to good marketing databases. On the contrary, the key question is having the necessary level of knowledge to take the right decisions (Campbell, 2003; Lin, Su, & Chien, 2006). The analytical capabilities of MkMSS are more critical than

ever to provide this support to marketing managers’ decision making, in order to give useful and valuable information about market behaviour. Specifically, we highlight the following: models and methods of analysis.

It is expected that MkMSS will improve their performance in the near future, taking advantage of the synergies caused by the integration of modelling estimation techniques based on classic econometrics with new methods and systems based on artificial intelligence (Gatignon, 2000; Van Bruggen & Wierenga, 2000). The adoption of these new methods represents a worthwhile opportunity to improve the efficiency of the marketing managers’ decision making and consequently, if well applied, the accuracy of marketing strategies (Li, Kinman, Duan, & Edwards, 2000).

The paper we present here focuses on the exploration and analysis of the suitability of certain brand new methods based on knowledge discovery in databases (KDD) to be applied in marketing modelling. Specifically, our main aim is twofold: first, we aim to make a modest contribution to the methods used in consumer behaviour modelling. In any case, this is the marketing field we have focused on to develop and experiment our methodology, though it also applies to marketing causal modelling, in general, as well as to other Science and Social Sciences fields that work with similar causal models.

We propose a complete knowledge discovery methodology, whose main questions are shown in this paper, to extract useful patterns of information with a descriptive rule induction approach based on Genetic Fuzzy Systems; this is a novel hybridization of methods belonging to the field of artificial intelligence, highly appropriate for the marketing problem we face. With this purpose, we have had to

* Corresponding author. Tel.: +34 958 242350.

E-mail addresses: fjmlopez@ugr.es (F.J. Martínez-López), casillas@decsai.ugr.es (J. Casillas).

¹ Tel.: +34 958 240804.

give solutions, adapted from our academic field, to the diverse questions related to the main stages of the KDD process; i.e. data preparation, data mining, and knowledge interpretation. Moreover, an important characteristic of our methodology is that it has been designed under the base there is a causal model of reference; a consumer behaviour model in our case. In other words, the knowledge discovery process is guided by a prior theoretic structure that defines the elements (variables) of the model. Hence, this machine learning approach is not only interesting for practitioners, but also for academic researchers' purposes.

To address these questions, the paper is structured as follows. Section 2 reflects on the suitability of evolving our marketing modelling methods towards a growing importation and use of artificial intelligence methods to support professional and academic marketing problems. Section 3 presents an overview and justification of the artificial intelligence tools employed (fuzzy rules, genetic algorithms, etc.). Section 4 illustrates with some examples the behaviour of the proposed KDD tools. Section 5 shows the methodological proposal in detail. Next, in Section 6 we experiment with the methodology, show some significant results and dedicate a brief closing part to illustrate both the intrinsic and complementary advantages of our fuzzy modelling-based method. Section 7 discusses the main contributions of our research, reflecting on the academic and managerial implications. Finally, in Section 8 we comment on some research limitations and opportunities (our future research agenda).

2. Background and starting reflections

Is there a gap between what marketing modellers offer and what marketing managers demand? If marketing modelling had got to a stage of maturity, as [Leeflang and Wittink \(2000\)](#) argue, one would expect to find a significant use of academic models among marketing practitioners. Notwithstanding, it seems that marketing managers rarely apply them ([Roberts, 2000](#); [Wind and Lilien, 1993](#); [Winer, 2000](#)). It is essential that we academics meditate on this. Maybe, the answer is much less complex than we would primarily expect.

We think that the efforts of marketing academics are not productive in terms of the managerial applications of their models. This is not due to deficiencies in the theoretic aspects that support the models' structure, but due to a lack of involvement by not offering useful methods of analysis that allow the models' users (marketing managers) to "play" with these models to support their decisions. This is what has guided our research, hence the gist of this paper.

The academics may be too focused on testing hypotheses and validating models and theories without paying enough attention to what our "customers"—the marketing managers, users of our scientific production—need. Indeed, marketing modellers cannot afford to fall into marketing myopia! In this regard, we should not forget that the main purpose of our research efforts ought to be the contribution to the development of our field, and this necessarily implies looking after the practical applicability of our models, too.

Therefore, how can we strengthen the utility of our models to achieve a better explanation of markets, thus better matching them to marketing managers' needs? Research efforts can be addressed to the improvement of three main areas of interest in marketing modelling ([Roberts, 2000](#)): theoretic aspects defining the models; understanding of managers' (users) needs, hence the framework of application of models; and refinement of the statistical tools (i.e. techniques and methods in general) applied to estimate the models. The pursuit of these improvement guidelines is not too distant from what [Little \(1970, p. B-483\)](#) asked of researchers a few decades ago when building models to support marketing managers' decision making:

Although the results of using a model may sometimes be personal to the manager [...] the researcher still has the responsibilities of a scientist in that he should offer the manager the best information

he can for making the model conform to reality in structure, parameterization, and behaviour.

Consequently, it seems clear that modellers should be driven by the requirements of models users (i.e. demand-side), instead of by a supply-side orientation ([Gatignon, 2000](#)). This practice is expected to improve the use of the academic models among the practitioners ([Roberts, 2000](#)). In this sense, a firm focused on consumption markets with access not only to more representative models of real systems being modelled but also to more powerful methods of analysis to extract knowledge from huge databases and able to simulate with models ought to improve its competitiveness and competitive advantage ([Van Bruggen & Wierenga, 2000](#)). This is a premise that has significantly conditioned the evolution of MkMSS from the early 80s, specifically with the appearance of the Marketing Decision Support Systems, until now ([Li et al., 2000](#); [Talvinen, 1995](#); [Wierenga & Van Bruggen, 1997, 2000](#)).

The late 80s saw the increasing use of diverse methods from Computer Science and Artificial Intelligence to the detriment of those from the Operational Research and, especially, the econometrics and statistics fields. This tendency has increasingly intensified in the last two decades ([Bucklin, Lehmann, & Little, 1998](#); [Eliashberg & Lilien, 1993](#); [Leeflang & Wittink, 2000](#); [Leeflang, Wittink, Wedel, & Naert, 2000](#); [Li, Davies, Edwards, Kinman, & Duan, 2002](#)).

This evolution in the methods used in marketing modelling has not been accidental. In this sense, [Lilien, Kotler, and Moorthy \(1992\)](#) noted that this tendency was to be expected as modellers and users needed techniques that were more flexible, powerful and robust, capable of providing greater and improved information with respect to the real systems being modelled. Of course, this implies a greater adaptation to both the characteristics of current databases—i.e. huge, imprecise, with data gathered in formats of a different nature (numerical, categorical, linguistic, etc.)—and the type of decision problems to be supported by such models. Under these circumstances, it seems an evolution of the marketing modelling methods towards systems based on artificial intelligence is only logical ([Shim et al., 2002](#); [Wedel, Kamakura, & Böckenholt, 2000](#)), which justifies the growing predominance of the knowledge-based MkMSS in the last two decades ([Wierenga & Van Bruggen, 2000](#)).

In sum, MkMSS clearly tend to be based on knowledge discovery methods that make use of diverse artificial intelligence methods to be applied during the machine learning process; e.g.: evolutionary algorithms, fuzzy logic, artificial neural networks, rules induction, decision trees, etc. Specifically, it is expected that the use of artificial intelligent methods in the MkMSS framework will evolve towards the use of intelligent systems based on the hybridization of these techniques ([Carlsson & Turban, 2002](#); [Shim et al., 2002](#)). We like to call them as Marketing Intelligent Systems. It might be the inexorable fate of marketing modelling methods. This fact, which is more evident from a professional perspective—i.e. under the framework of application of the MkMSS—, has still to take hold in academic studies.

3. Knowledge extraction based on fuzzy rules and genetic algorithms

3.1. The KDD process

In general terms, KDD is a recent research field belonging to artificial intelligence whose main aim is the identification of new, potentially useful, and understandable patterns in data ([Fayyad, Piatesky-Shapiro, Smyth, & Uthurusamy, 1996](#)). Furthermore, KDD implies the development of a process compounded by several stages that allow the conversion of low-level data into high-level knowledge ([Mitra, 2002](#)). Though KDD is synthetically viewed as a three-stage process—i.e. pre-processing, data mining and post-processing—([Freitas, 2002](#)), we believe that, for our academic field, it is more interesting to present it within a wider structure. Specifically, we prefer the following five-stage process

(Cabena, Hadjinian, Stadler, Verhees, & Zanasi, 1998; Han & Kamber, 2001): (1) identification and problem delimitation; (2) data preparation (pre-processing); (3) data mining (machine learning); (4) analysis, evaluation and interpretation of results; and (5) presentation, assimilation and use of knowledge. It is important to highlight that the success of such process, applied to solve or support the resolution of a particular problem of information in marketing, depends on the suitable development of every stage. The reader will be more conscious of this question when observing the lengths we go in order to explain how to prepare marketing data (pre-processing) or how to analyse the output (knowledge) of the data mining stage (post-processing).

3.2. Knowledge representation by fuzzy rules

Nowadays, one of the most successful tools for the development of descriptive models is fuzzy modelling (Lindskog, 1997). This is an approach used to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates. The way to express fuzzy predicates is by means of IF-THEN rules, as in the following example:

IF Age_of_Consumer is *Young* and Purchasing_Power is *Very_High*
THEN Trend_To_Buy_Sports_Cars is *High*

These rules set logical relationships among variables of a system by using qualitative values. Such a representation mode easily matches the humans' way of reasoning. Hence, the performance of both the analysis and interpretation steps of the modelling process improves thanks to the true behaviour of a system that is more effectively revealed. Notwithstanding, it should be noted that though human reasoning may deal without difficulty with terms like *high* or *young*, when this issue is tackled by means of an automatic process its treatment is more complex.

To work properly with these kinds of qualitative valuations, linguistic variables (Zadeh, 1975a,b, 1976) based on both Fuzzy Sets Theory and Fuzzy Logic (Zadeh, 1965) are used, so the previously exemplified rule is known as a *fuzzy rule*. The use of fuzzy logic provides several benefits, such as a higher generality, expressive power, ability to model real problems and, last but not least, a methodology to exploit tolerance in the face of imprecision.

For example, we can consider the linguistic variable *Age_of_Consumer*, which could take in the linguistic terms (values) *teenager*, *young*, *adult*, and *old*. These linguistic terms (also known as labels) are mathematically expressed by simple functions that return the membership degree (with a real value between 0 and 1) to each fuzzy set. Therefore, instead of considering that a consumer could be 100% young or 100% adult, with fuzzy sets we can say that the consumer belongs to the set of young people with one degree and also to the set of adults with another degree. So, the boundaries between sets are fuzzy instead of crisp, thus providing a powerful linguistic expression and a gradual transition of the membership to the different fuzzy sets.

Fig. 1 represents an example of how the age of a person can be expressed by fuzzy sets. In this figure, we could say that a person of

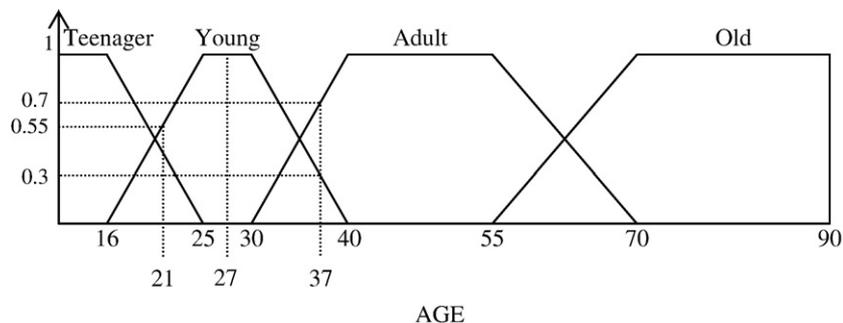


Fig. 1. Illustrative example of the linguistic variable *age*, composed of the linguistic terms *teenager*, *young*, *adult* and *old*, and their corresponding fuzzy sets. A 37-year-old has a membership degree 0.3 to *young* and 0.7 to *adult*.

21 years of age belongs to the fuzzy set labelled *young* to a degree of 0.55 (colloquially speaking, 55%), while a 27 year-old belongs entirely to the fuzzy set *young*, or a 37-year-old belongs to *young* people to a 0.3 degree and also to *adult* people to a degree of 0.7. If we used classical (crisp) sets and fixed the boundary between *young* and *adult* at 35 years of age, a person aged 34.9 years would be considered 100% *young* while another aged 35.1 years would not be *young* to any degree.

Fuzzy rules can be considered a useful representation of knowledge to discover intrinsic relationships contained in a database (Freitas, 2002). Thus, by means of fuzzy rules we can represent the relationship existing among different variables, thus deducing the patterns contained in the data examined. Useful patterns allow us to do non-trivial predictions about new data. There are two extremes to express a pattern: *black boxes*, whose internal behaviour is incomprehensible; and *white boxes*, whose construction reveals the pattern structure. The difference lies in whether the patterns generated are represented by a structure that is easy to examine and which can be used to reason and to inform further decisions. In other words, when the patterns are structured in a comprehensible way, they will be able to help explain something about the data. The trouble with KDD, the interpretability-accuracy trade-off, is also being tackled in current fuzzy modelling (Casillas et al., 2003a,b) and will be considered by our proposal.

The use of fuzzy rules when developing the knowledge discovery process has some advantages, which are (Freitas, 2002; Dubois, Prade, & Sudkamp, 2005): they allow us to deal with uncertain data; they adequately consider multi-variable relationships; results are easily understandable by humans; additional information is easily added by an expert; the accuracy degrees can be easily adapted to the needs of the problem, and the process can be highly automatic with low human intervention.

Therefore, we will use fuzzy logic as a tool to structure the information of a consumer behaviour model in a clear and intelligible way that is close to that of the human being. Fuzzy logic methods are expected to offer benefits to marketing decision makers when integrated with current MkMSS (Metaxiotis, Psarras, & Samouilidis, 2004). The fuzzy system will allow us to represent adequately the interdependence of variables and the non-linear relationships that could exist between them.

3.3. Multiobjective genetic algorithms

In the previous section, we introduced the proposed representation of knowledge based on fuzzy rules. However, we also need an algorithm to automatically extract a set of fuzzy rules with good properties. In this paper, we propose the use of a genetic algorithm. The main reasons for using it instead of other well-known machine learning techniques are the following. Firstly, since there are usually contradictory objectives to be optimised in KDD (such as accuracy and interpretability, or support and confidence), we perform multiobjective optimisation. It is one of the most promising issues and one of the main distinguishing features of genetic algorithms compared to other techniques. Furthermore, we consider a flexible representation of fuzzy rules that can be developed

```

Genetic Algorithm
begin
  t = 0
  initialize P(t)
  evaluate P(t)
  while (not termination-condition) do
    t = t+1
    P(t) ← select P(t-1)
    P(t) ← crossover P(t)
    P(t) ← mutation P(t)
    evaluate P(t)
  end-while
end

```

Fig. 2. Structure of a genetic algorithm.

properly by genetic algorithms. This flexible representation improves the description capability of the fuzzy rule, an important issue in KDD.

Genetic algorithms demonstrated good results for management and marketing applications, thus arousing the interest of researchers and practitioners in the nineties (Hurley, Moutinho, & Stephens, 1995; Nissen, 1995). However, one of the novelties of this paper for marketing is that, in this instance, fuzzy logic and genetic algorithms will not be applied separately to tackle a particular marketing problem, but in cooperation. In the following, genetic algorithms and multiobjective optimisation are briefly introduced.

3.3.1. Genetic algorithms

Genetic algorithms are general-purpose search algorithms that use principles inspired by natural population genetics to evolve solutions to problems. The basic principles of genetic algorithms were first laid down rigorously by Holland (1975) and are well described in many texts (e.g.: Goldberg, 1989; Michalewicz, 1996).

The basic idea is to maintain a population (i.e., a set) of knowledge structures that evolves over time through a process of competition and controlled variation. Each structure in the population represents a candidate solution to the specific problem and has an associated *fitness* to determine which structures are used to form new ones in the process of competition. The new individuals are created using genetic operators such as crossover and mutation. Genetic algorithms have had a great measure of success in search and optimisation problems. The main reason for this success is their ability to exploit accumulative information about an initially unknown search space in order to bias subsequent search into useful subspaces, i.e., *their robustness*. This is their key feature, especially in large, complex and poorly understood search spaces, where the classical search tools (enumerative, heuristic, etc.) are inappropriate, offering a valid approach to problems requiring efficient and effective search.

A genetic algorithm starts with a population of randomly generated solutions, chromosomes, and advances towards better solutions by applying genetic operators, modelled on the genetic processes occurring in nature. As previously mentioned, in these algorithms we maintain a population of solutions (in our case, fuzzy rules) for a given problem; this population undergoes evolution in a form of natural selection. In each generation, relatively good solutions reproduce to give offspring that replace the relatively bad solutions, which die. An evaluation or fitness function plays the role of the environment to distinguish between good and bad solutions. The process of evolving from the current population to the next one constitutes one generation in the execution of a genetic algorithm.

Although there are many possible variants of the basic genetic algorithm, the fundamental underlying mechanism involves three operations (Goldberg, 1989):

- (1) evaluation of individual fitness,
- (2) formation of a gene pool (intermediate population), and
- (3) crossover and mutation.

Fig. 2 shows the structure of a simple GA.

A fitness function must be devised for each problem to be solved. Given a particular chromosome (i.e. a solution), the fitness function returns a numerical value that is supposed to be proportional to the utility or adaptation of the solution represented by this chromosome. In our case, we will consider two different measures to assess the quality of a solution (fuzzy rule): support and confidence.

There are a number of ways to do selection. We might view the population as a mapping on a roulette wheel, where each individual is represented by a space that proportionally corresponds to its fitness. By repeatedly spinning the roulette wheel, individuals are chosen using stochastic sampling with replacement to fill the intermediate population. Another possibility, called binary tournament, consists in doing a number of tournaments equal to the size of the population. In each tournament, two chromosomes of the old population are chosen at random, and the best one according to fitness is included in the new population. We will employ this second approach in our proposal.

After selection has been carried out, the construction of the intermediate population is completed and crossover and mutation can occur. The crossover operator combines the features of two parent structures to form two similar offspring. Classically, it is applied at a random position with a probability of performance, the crossover probability. The mutation operator arbitrarily alters one or more components of a selected structure so as to increase the structural variability of the population. Each position of each solution vector in the population undergoes a random change according to a probability defined by a mutation rate, the mutation probability.

Fig. 6 in Section 4 illustrates graphically the use of a genetic algorithm to extract fuzzy rules from available data in the marketing problem we are dealing with in this paper.

3.3.2. Multiobjective optimisation

Many real-world problems involve simultaneous optimisation of multiple objectives. In principle, multiobjective optimisation is very different from single-objective optimisation. The second case attempts to obtain the best solution; i.e. the global minimum or the global maximum depending on the problem. However, in the case of multiple objectives, there may not be a single solution that is better than the rest with respect to all objectives.

In a typical multiobjective optimisation problem, there is a set of solutions that are superior to the rest of the solutions in the search space when all the objectives are considered, but which are inferior to other solutions in the space occupied only by some of them. These solutions are known as *non-dominated solutions* (Chankong & Haimes, 1983), while the rest of the solutions are known as *dominated solutions*. Since none of the solutions in the non-dominated set is worse in all the objectives than the other ones, all of them are acceptable solutions.

Mathematically, the concept of *Pareto-optimality*² or *non-dominance* is defined as follows. Let us consider, without loss of generality, a multiobjective maximization problem with m parameters (decision variables) and n objectives:

$$\text{Maximise } f(x) = (f_1(x), f_2(x), \dots, f_n(x))$$

with $x = (x_1, x_2, \dots, x_m) \in X$. A decision vector $a \in X$ dominates $b \in X$ (noted as $a \leq b$) if, and only if:

$$\forall i \in \{1, \dots, n\} | f_i(a) \geq f_i(b) \text{ and } \exists j \in \{1, \dots, n\} | f_j(a) > f_j(b).$$

Any vector that is not dominated by any other is said to be *Pareto-optimal* or *non-dominated*. These concepts are depicted graphically in Fig. 3.

² The concept *Pareto optimality* is an important notion in neoclassical economics. It is named after the French-Italian economist Vilfredo Pareto (1848, Paris–1923, Geneva).

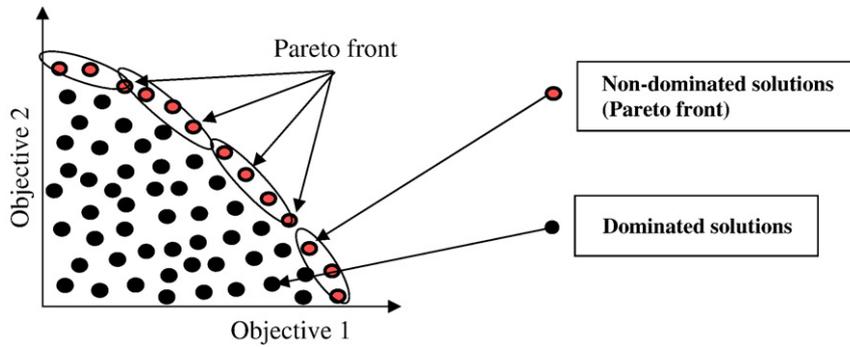


Fig. 3. Example of multiobjective optimisation.

Thanks to the use of a solution population, genetic algorithms can simultaneously search for many Pareto-optimum solutions. For this reason, genetic algorithms have been recognised as possibly being well suited to multiobjective optimisation (Coello, Van Veldhuizen, & Lamont, 2002). Furthermore, the ability to handle complex problems, involving

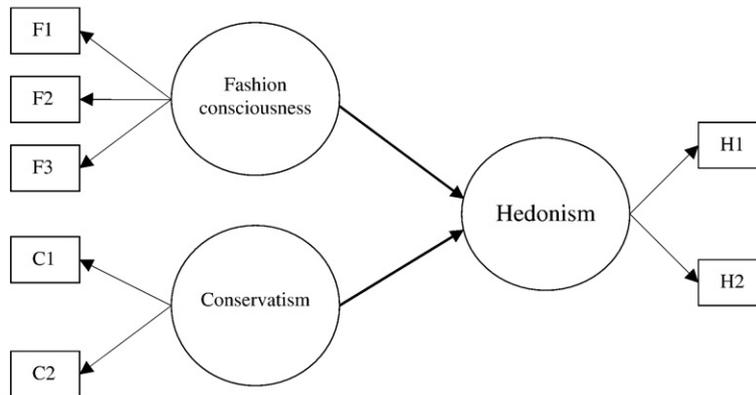
features such as discontinuities, multimodality, disjoint feasible spaces, and noisy function evaluations, reinforces the potential effectiveness of genetic algorithms in multiobjective search and optimisation. Generally, the multiobjective approaches only differ from the rest of the genetic algorithms in the fitness function and/or in the selection operator.

4. An illustrative example on how to extract knowledge from data to analyse consumer behaviour

This section serves as a bridge between the technical concepts included in the previous section and the modelling methodology proposed in the next one. Therefore, to introduce the reader to the methodology, we propose extracting useful knowledge from data that can aid better understanding the existing relationships between variables by presenting in this section a toy problem (with a few variables and a small data set size) to illustrate the basic behaviour and powerful nature of the proposed KDD process. Some parts of the process have been intentionally simplified with the aim of focusing on the most relevant aspects. The rigorous description of the proposal can be found in Section 5, while Section 6 amply describes the experimental results in a real-world problem.

To illustrate the proposed use of KDD, we will consider a simple measurement (causal) model depicted in Fig. 4(a), compounded by three construct or latent variables (depicted by circles), two exogenous and one endogenous: (1) *fashion consciousness*, (2) *conservatism*, and (3) *hedonism*; extracted from MacLean and Gray (1998). Likewise, imagine that the three constructs have been measured by means of several seven-point interval scales (e.g. Likert-type and differential semantic scales). Finally, Fig. 4(b) shows an example of a data set available for this problem, which consists of three variables, each made up of a set of values. There are just four cases (e.g., questionnaires), which are not realistic at

(a) Measurement (causal) model



(b) Example of a data set from four responses about the items considered in the above model

Cases	Fashion Consciousness			Conservatism		Hedonism	
	F1	F2	F3	C1	C2	H1	H2
Consumer 1	1	2	1	5	6	1	2
Consumer 2	5	5	6	2	1	7	6
Consumer 3	7	7	7	1	2	7	7
Consumer 4	4	4	4	3	3	3	2

Fig. 4. Example of a simple measurement (causal) model—extracted from (MacLean & Gray, 1998)—and a data set from four hypothetical consumers' responses.

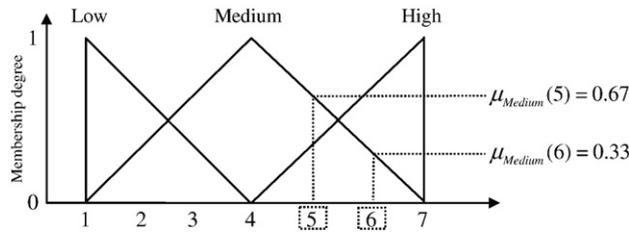


Fig. 5. Example of transformation of a seven-point Likert-type scale into a fuzzy semantic. According to that, the membership degree of 5 to the fuzzy set associated to the linguistic term *Medium* is 0.67, while the membership degree of 6 is 0.33.

all–i.e. think that a consumer database usually has hundreds or even thousands of collected individuals' responses–, though it is useful for our illustrative purpose.

The first step we perform is to transform the interval scale into fuzzy semantics. This allows us to use linguistic terms to describe the different items by means of linguistic variables. We can consider the following three membership functions to describe the terms *Low*, *Medium*, and *High*:

$$\mu_{Low}(x) = \begin{cases} \frac{4-x}{3} & \text{if } 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}, \quad \mu_{Medium}(x) = \begin{cases} \frac{x-1}{3} & \text{if } 1 \leq x \leq 4 \\ \frac{7-x}{3} & \text{if } 4 < x \leq 7 \\ 0 & \text{otherwise} \end{cases}, \quad \mu_{High}(x) = \begin{cases} \frac{7-x}{3} & \text{if } 5 \leq x \leq 7 \\ 0 & \text{otherwise} \end{cases}$$

A graphical representation of these membership functions is depicted in Fig. 5.

Once we have defined the variables in terms of fuzzy sets, we can use fuzzy rules to express relationships (i.e., patterns) among the variables (refer to Section 3.2. for a description of these kinds of rules). To do that, we will consider the two exogenous variables and the endogenous one, antecedents and consequent respectively in this example.

These fuzzy rules can represent many different relationships among the variables; however, not all of them will match the existing data exactly. Therefore, we need some measures to assess the quality of each rule with respect to the data. These measures can be considered a kind of statistical computation. In this paper, we will consider two important values: support and confidence. On the one hand, *support* (whose real value is in [0,1]) will give us an idea about in which degree the rule represents the cases of the data set. For example, a support of 0.25 could be understood as the rule that covers 25% of the available cases. We are interested in obtaining fuzzy rules with a support as high as possible since the rule will be more general and will represent a higher portion of the sample. On the other hand, *confidence* (whose real value is also in [0,1]), indicates how accurate the fuzzy rule is. Since the fuzzy rule predicts a relationship between the antecedent and the consequent, we need to know in which degree such a prediction appears in the available data set. For example, if a fuzzy rule has a confidence of 0.9, we can say that, according to the available data, the fuzzy rule is 90% true. Of course, we are interested in obtaining fuzzy rules with a high degree of confidence.

As one can imagine, support and confidence are two contradictory features. Inasmuch as the degree of representation is higher, it is more difficult to accurately express the relationships among variables. One fuzzy rule will be clearly preferable to another if the former has higher values of both support and confidence.

In the following, we will show some examples of fuzzy rules and the computation of the corresponding support and confidence values from the data set of Fig. 4(b).

R₁: If Fashion_Consciousness is LOW and Conservatism is MEDIUM then Hedonism is MEDIUM

$$\mu_{Low}(\vec{x}_1^{(1)}) = \max \{ \mu_{Low}(1), \mu_{Low}(2), \mu_{Low}(1) \} = \max \{ 1, 0.67, 1 \} = 1$$

$$\mu_{Medium}(\vec{x}_2^{(1)}) = \max \{ \mu_{Medium}(5), \mu_{Medium}(6) \} = \max \{ 0.67, 0.33 \} = 0.67$$

$$\mu_{A^{(1)}}(x^{(1)}) = \min \{ \mu_{Low}(\vec{x}_1^{(1)}), \mu_{Medium}(\vec{x}_2^{(1)}) \} = \min \{ 1, 0.67 \} = 0.67$$

$$\mu_{Medium}(\vec{y}^{(1)}) = \max \{ \mu_{Medium}(1), \mu_{Medium}(2) \} = \max \{ 0, 0.33 \} = 0.33$$

$$\mu_{Low}(\vec{x}_1^{(2)}) = 0, \mu_{Medium}(\vec{x}_2^{(2)}) = 0.33, \mu_{A^{(1)}}(x^{(2)}) = 0, \mu_{Medium}(\vec{y}^{(2)}) = 0.33$$

$$\mu_{Low}(\vec{x}_1^{(3)}) = 0, \mu_{Medium}(\vec{x}_2^{(3)}) = 0.33, \mu_{A^{(1)}}(x^{(3)}) = 0, \mu_{Medium}(\vec{y}^{(3)}) = 0$$

$$\mu_{Low}(\vec{x}_1^{(4)}) = 0, \mu_{Medium}(\vec{x}_2^{(4)}) = 0.67, \mu_{A^{(1)}}(x^{(4)}) = 0, \mu_{Medium}(\vec{y}^{(4)}) = 0.67$$

$$\text{Support}(R_1) = \frac{1}{4} \sum_{e=1}^4 (\mu_{A^{(1)}}(x^{(e)}) \cdot \mu_{B^{(1)}}(\vec{y}^{(e)})) = \frac{0.67 \cdot 0.33 + 0 + 0 + 0}{4} = 0.05556$$

$$\text{Confidence}(R_1) = \frac{\sum_{e=1}^4 (\mu_{A^{(1)}}(x^{(e)}) \cdot \max \{ 1 - \mu_{A^{(1)}}(x^{(e)}), \mu_{B^{(1)}}(\vec{y}^{(e)}) \})}{\sum_{e=1}^4 \mu_{A^{(1)}}(x^{(e)})} = \frac{(0.67 \cdot \max \{ 1 - 0.67, 0.33 \} + 0 + 0 + 0)}{0.67 + 0 + 0 + 0} = 0.33333$$

R_2 : If Fashion_Consciousness is MEDIUM and Conservatism is MEDIUM then Hedonism is MEDIUM

$$\begin{aligned} \mu_{\text{Medium}}(\vec{x}_1^{(1)}) &= 0.33, \mu_{A(2)}(x^{(1)}) = 0.33 & \mu_{\text{Medium}}(\vec{x}_1^{(2)}) &= 0, \mu_{A(2)}(x^{(2)}) = 0 \\ \mu_{\text{Medium}}(\vec{x}_1^{(3)}) &= 0, \mu_{A(2)}(x^{(3)}) = 0 & \mu_{\text{Medium}}(\vec{x}_1^{(4)}) &= 1, \mu_{A(2)}(x^{(4)}) = 0.67 \end{aligned}$$

$$\text{Support}(R_2) = \frac{0.33 \cdot 0.33 + 0 + 0 + 0.67 \cdot 0.67}{4} = 0.13889$$

$$\text{Confidence}(R_2) = \frac{0.33 \cdot \max\{1 - 0.33, 0.33\} + 0 + 0 + 0.67 \cdot \max\{1 - 0.67, 0.67\}}{0.33 + 0 + 0 + 0.67} = 0.44445.$$

As we can observe, the fact of using the linguistic term “medium” for the fashion consciousness variable instead of “low” (as in rule R_1) allows us to cover better the data set and, at the same time, to improve the accuracy of the rule.

R_3 : If Fashion_Consciousness is MEDIUM and Conservatism is {LOW or MEDIUM} then Hedonism is MEDIUM

$$\mu_{\text{Low or Medium}}(\vec{x}_2^{(1)}) = \min\{1, \mu_{\text{Low}}(\vec{x}_2^{(1)}) + \mu_{\text{Medium}}(\vec{x}_2^{(1)})\} = 0.67, \mu_{A(3)}(x^{(1)}) = 0.33$$

$$\mu_{\text{Low or Medium}}(\vec{x}_2^{(2)}) = 1, \mu_{A(3)}(x^{(2)}) = 0.33$$

$$\mu_{\text{Low or Medium}}(\vec{x}_2^{(3)}) = 1, \mu_{A(3)}(x^{(3)}) = 0.33$$

$$\mu_{\text{Low or Medium}}(\vec{x}_2^{(4)}) = 1, \mu_{A(3)}(x^{(4)}) = 0.67$$

$$\text{Support}(R_3) = 0.16667 \quad \text{Confidence}(R_3) = 0.66667.$$

This third rule includes two linguistic terms in the variable conservatism. Doing that, the support is higher since we can cover the data set to a higher degree compared to rule R_2 (it is obvious since R_3 is more general than R_2). Moreover, the confidence is also improved, so this third rule is clearly better than the previous ones.

Once we have shown some examples of fuzzy rules and how to compute their associated support and confidence values from a data set, we will illustrate a simplification of how the data mining process works. Fig. 6 depicts a scheme of the behaviour of a genetic algorithm to reveal fuzzy rules from data. The genetic algorithm, as explained in Section 3.3.1, optimises generation by generation the population, in our case a set of different fuzzy rules, i.e., patterns. To analyse alternative fuzzy rules, new ones are generated from the existing one by applying the crossover and mutation operators. The genetic algorithm encodes the rules in a format that is easily tractable in a computer, in this case by using a binary representation.

In the example of Fig. 6, the mutation takes a solution from the current population and applies a slight alteration; in this case, it changes the linguistic term used in the first variable from “low” to “medium.” The new generated rule is included in the next population since its corresponding values of support and confidence are better. In other example, the crossover takes two solutions and combines them by generating a new rule that contains the linguistic terms considered in each parent rule. This new rule, better than its parents, is included in the new population.

5. A marketing intelligent system for consumer behaviour analysis

This section introduces the process in which we propose performing knowledge discovery related to consumers by fuzzy rules. Basically, it consists of preparing the data and of fixing the scheme we follow to represent the knowledge existing in the data. Once these aspects are defined, a machine learning method is used to automatically extract interesting fuzzy rules. Finally, a post-processing stage is carried out. All these questions are now presented in detail.

5.1. Data gathering

First step is to collect the data related to the variables defining the theoretical model of the consumer behaviour proposed. In this sense, as has been done traditionally in Marketing Science in particular, and in Social Sciences in general, data is obtained by means of a questionnaire. This questionnaire gathers the measures for the set of constituent elements of the model.

5.2. Data processing

Next, it is necessary to adapt the collected data to a scheme easily tractable by fuzzy rule learning methods. Thus, at first, attention should be paid to how modellers face and develop the measurement process of the elements/variables contained in the complex behavioural models. In this respect, reflections about the measurement of such variables, with a special focus on those usually known as

theoretical constructs (i.e. unobserved variables), should be made. Consequently, we think that time should be spent analysing the adaptation of the fuzzy rule-based KDD to the latter case, inasmuch as its treatment seems to be the more controversial.

Previously, it could be said that measuring streams for these latent variables in consumer modelling was classified into two groups depending on if they declared that these constructs could or could not be perfectly measured by means of observed variables (indicators): the *operational definition philosophy* and the *partial interpretation philosophy* respectively. This latter approach of measurement, currently predominant in the marketing modelling discipline, recognises the impossibility of doing perfect measurements of theoretical constructs by means of indicators, so it poses joint consideration of multiple indicators—imperfect when considered individually, though reliable when considered together—of the subjacent construct to obtain valid measures (Steenkamp & Baumgartner, 2000).

Therefore, our methodological approach should be aware of this question when adapting the data (observed variables) to a fuzzy rule learning method. Notwithstanding, we would like to highlight that our method does not have any problem with processing elements of a model for which we have just a single variable or indicator associated to each of them, even when they have been measured by varied measurement scales. The problem comes, hence the challenge to face, when there are multiple variables related to the measurement of a particular element of the model. Some intuitive solutions and aprioristic analyses of the internal consistency of the multi-item scales associated to such elements have been proposed, with the aim of keeping just

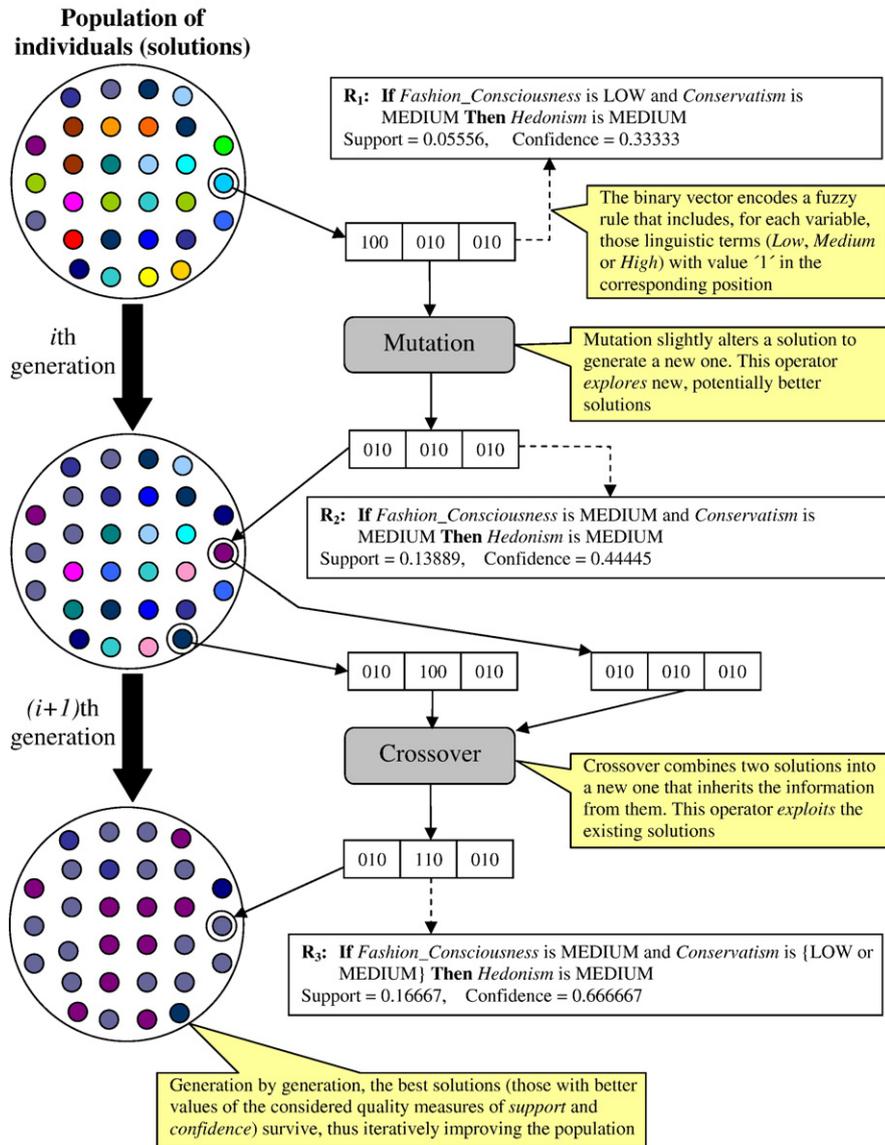


Fig. 6. A simplified example of the behaviour of a genetic algorithm when extracting knowledge in form of fuzzy rules from the data set available in Fig. 4(b).

one indicator (the best) per construct (see: Casillas, Martínez-López, & Martínez, 2004). The weakness of these approaches is that the data must be transformed, so relevant information may be lost.

We propose a solution based on a more sophisticated process that allows working with the original format without any pre-processing stage (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, can be applied to aggregate the partial information given by each item. Since it is not pre-processing data but a component of the machine learning design, the details of that treatment of the items is described in Section 5.4.2.

5.3. Representation and inclusion of the marketing expert's knowledge

Several issues should be tackled at this step of our methodological proposal: the set of variables/constructs to be processed, the transformation of the marketing scales used for measuring such variables into fuzzy semantic, the relations among constructs (i.e. the causal model), and the fuzzy rules' sets to be generated. All of them are based on the expert's capability to express his knowledge in a humanly understandable format by fuzzy logic.

5.3.1. Fuzzy semantics from expert knowledge

Once the marketing modeller has finally determined both the elements of the model and the observed variables associated to each one (i.e. the measurement model), a transformation into linguistic terms (fuzzy semantic) of the original marketing scales used for measuring those observed variables should be done. This is necessary for the derivation of fuzzy rules later. This question implies treating the application of the fuzzy set theory to the measurement in marketing. In this regard, as far as we know, Viswanathan, Bergen, Dutta, and Childers (1996) were the ones who first researched this question by proposing a methodology for the scale development in marketing. In any case, as this is not the central theme of this paper, we are not going to treat this issue in depth, though it is thoroughly analysed in the research that supports this study.

Several marketing scale types can be used to measure the variables associated to the constituent elements of a consumer behaviour model. With the aim of focusing the problem, we take Stevens (1946, 1959) as a base to summarize them in four categories with regard to their level of measurement, i.e. nominal, ordinal, interval and ratio. Considering those types, a transformation into fuzzy semantic is meaningful for the majority with the exception of variables measured by means of a nominal scale, where the nature of categories defining the scale

are purely deterministic. In general terms, this transformation should be practiced taking into account two main questions:

- The *number of linguistic terms* to be used, which determines the granularity (the scale sensitivity) of certain fuzzy variable, must be defined. Thus, although more terms are used, the analysis of relations among variables is more accurate, but more complex too. Consequently, the marketing modeller should take time to think about what the most convenient degree of sensitivity is in the fuzzy scales used in his/her study. Three or five linguistic terms (fuzzy sets) seem good options.
- The *membership function type and shapes* defining the behaviour of a certain fuzzy variable should be also defined. Such behaviour can be broadly treated considering the use of linear vs. non-linear membership functions to characterise the fuzzy sets. Thus, trapezoidal and triangular functions can be used to obtain a linear behaviour, while Gaussian functions can be used for a non-linear one.

We are now going to focus on those marketing scales mainly used for measuring the observed variables related to the elements (theoretical constructs) of a particular marketing model; i.e.: Likert-type and differential semantic. Firstly, we have considered that it is more appropriate to use linear functions, inasmuch as it facilitates the interpretation of relations later. Second, we believe that a transformation into a triangular function is more convenient if special characteristics of these marketing scales are considered; scales valuations are punctual. Then, when the membership degree of certain linguistic terms is equal to one, such a term should be associated to a point of the scale. In this regard, this choice has also been justified in the marketing context, with the argument that trapezoidal functions facilitate the later process of fuzzy inference (Li et al., 2002).

To sum up, Fig. 5 shows an example based on the transformation of a seven-point rating scale into a three-triangular fuzzy semantic, with the three linguistic terms (Low, Medium, and High) represented by the corresponding fuzzy sets characterised by the three membership functions shown in Section 4.

5.3.2. Input/output linguistic variables from expert knowledge

Once the causal model has been fixed by the marketing expert, fuzzy rules are used to relate input (antecedents) with output (consequents) variables. Obviously, the theoretic relations defining the model can be directly used to define the IF-THEN structures by considering the dependences shown among the variables. Thus, we obtain a set of fuzzy rules for each considered consequent (i.e. endogenous element of the model) and its respective set of antecedents. Several examples of fuzzy rules from the model included in Fig. 4(a) can be found in Section 4.

5.4. Machine learning (data mining process)

5.4.1. Fuzzy rule structure

In data mining, it is crucial to use a learning process with a high degree of interpretability preservation. To do that, we can opt for using a compact description as the *disjunctive normal form*. This kind of fuzzy rule structure has the following form (González & Pérez, 1998):

R: IF X_1 is \tilde{A}_1 and ... and X_n is \tilde{A}_n THEN Y_1 is B_1 and ... Y_m is B_m with each input variable X_i , $i \in \{1, \dots, n\}$, taking as a value a set of linguistic terms $\tilde{A}_i = \{A_{i1} \text{ or } \dots \text{ or } A_{im}\}$, whose members are joined by a disjunctive (T-conorm) operator, while the output variables Y_j , $j \in \{1, \dots, m\}$, remain a usual linguistic variable with single labels associated. We use the *bounded sum* as T-conorm in this paper:

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

This structure uses a more compact description that improves interpretability. Moreover, the structure is a natural support to allow for the absence of some input variables in each rule (simply making \tilde{A}_i to be the whole set of linguistic terms available).

5.4.2. Multi-item fuzzification

In order to properly consider the set of items available for each input/output variable (as discussed in Section 5.2), 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 $\vec{x}_i = (x_1^{(i)}, \dots, x_{h_i}^{(i)}, \dots, x_{p_i}^{(i)})$ and $\vec{y}_j = (y_1^{(j)}, \dots, y_{t_j}^{(j)}, \dots, y_{q_j}^{(j)})$, respectively, the fuzzy propositions “ X_i is \tilde{A}_i ” and “ Y_j is B_j ” are respectively interpreted as follows:

$$\mu_{\tilde{A}_i}(\vec{x}_i) = \max_{h_i=1}^{p_i} \mu_{\tilde{A}_i}(x_{h_i}^{(i)}) \quad \text{and} \quad \mu_{B_j}(\vec{y}_j) = \max_{t_j=1}^{q_j} \mu_{B_j}(y_{t_j}^{(j)}).$$

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

5.4.3. Discovery process

In order to perform descriptive induction we will apply a method with some similarities to subgroup discovery, widely used in learning classification rules (Lavrac, Cestnik, Gamberger, & Flach, 2004) where the interest property is the class associated to the consequent variable. Therefore, this technique seeks to group the set of data into different subgroups, including in each of them the example set by the corresponding consequent, and to discover a set of rules representing this subgroup. In that case, the most usual approach involves running the algorithm once for each subset of examples holding the property fixed for the consequent.

Instead of that, our algorithm considers the subgroup division according to the used fuzzy set in the consequent; therefore, the subsets of examples can be overlapped. Moreover, we propose performing a *simultaneous* subgroup discovery where niches of fuzzy rules, in accordance with the consequent, are formed and optimised in parallel to generate a final set of suboptimal solutions in each subgroup. To perform this process, as explained in the following sections, we vary the concept of multiobjective dominance and we design the genetic operators for acting only on the antecedent part.

5.4.4. Coding scheme

Each individual of the population represents a fuzzy rule. The rule 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].

5.4.5. Objective functions

We consider the two criteria most often used to assess the quality of association rules (Dubois et al., 2005): support and confidence. In Section 4, the reader can see some examples of how these measures are computed.

- Support*: This objective function measures the representation degree of the corresponding fuzzy rule among the available data. It is computed as the mean covering degree of the rule for each data. As covering, we consider the conjunction of the

membership degrees of both antecedent and consequent variables. Therefore, the support measure (for maximization) of the fuzzy rule R: $A \Rightarrow B$ is defined as follows:

$$\text{Sup}(R) = \frac{1}{N} \sum_{e=1}^N \mu_A(\mathbf{x}^{(e)}) \cdot \mu_B(\bar{\mathbf{y}}^{(e)})$$

with N being the data set size, $\mathbf{x}^{(e)} = (\bar{x}_1^{(e)}, \dots, \bar{x}_n^{(e)})$ and $\bar{\mathbf{y}}_e$ the e th input/output multi-item data instance, and $\mu_A(\mathbf{x}^{(e)}) = \min_{i \in \{1, \dots, n\}} \mu_{\bar{A}_i}(\bar{x}_i^{(e)})$ the covering degree of the antecedent of the rule R for each example (i.e., the T-norm *minimum* is considered to interpret the connective ‘and’ of the fuzzy rule). As shown, the T-norm of the *product* is considered as joint antecedent and consequent. Note that we use the multi-item fuzzification described in Section 5.4.2 to compute $\mu_{\bar{A}_i}(\bar{x}_i^{(e)})$ and $\mu_B(\bar{\mathbf{y}}^{(e)})$.

(2) *Confidence*: This second objective measures the reliability of the relation between antecedent and consequent described by the analysed fuzzy rule. We have used a confidence measure that avoids the accumulation of low cardinalities (Dubois et al., 2005). It is computed (for maximization) as following:

$$\text{Conf}(R) = \frac{\sum_{e=1}^N (\mu_A(\mathbf{x}^{(e)}) \cdot \max\{1 - \mu_A(\mathbf{x}^{(e)}), \mu_B(\bar{\mathbf{y}}^{(e)})\})}{\sum_{e=1}^N \mu_A(\mathbf{x}^{(e)})}$$

Therefore, the Dienes’ S-implication, $I(a,b) = \max\{1 - a, b\}$, is used. Note that this implication operator is a fuzzy interpretation of the classical interpretation $p \Rightarrow q \equiv \neg p \vee q$ used in Boolean logic where the negation is interpreted as $1 - a$ and the disjunction as $\max\{a, b\}$. Multi-item fuzzification is again considered.

5.4.6. Evolutionary scheme

We consider a generational approach with the multiobjective elitist replacement strategy of NSGA-II (Deb, Pratap, Agarwal, & Meyarevian, 2002). Crowding distance in the objective function space is used. Binary tournament selection based on the non-domination rank (or the crowding distance when both solutions belong to the same front) is applied.

To perform simultaneous subgroup discovery properly, we need to redefine the dominance concept. Thus, one solution (fuzzy rule) dominates another when, besides being better or equal in all the objectives and better in at least one of them, it has the same consequent as the other rule. In that way, those rules with different consequents are not dominated between them, thus inducing the algorithm to form a search niche (Pareto set) for each considered consequent (subgroup).

5.4.7. Genetic operators

The *initial population* is built by defining the same amount of groups (with the same size) as the consequents considered. In each of them, the chromosomes are generated by fixing the consequent and by randomly defining a simple antecedent to which each variable is assigned only one linguistic term. The two genetic operators (crossover and mutation) act only on the antecedent part. This allows the algorithm to keep a constant size for each subgroup.

The *crossover* operator randomly chooses two cross points (in the antecedent) and exchanges the central string of the two selected parents. If all the linguistic terms of a variable are set off after crossover, a linguistic term used in the parents is randomly chosen and set to ‘1’. It is interesting to note that no constraints are imposed on selecting the parents, so the crossover can be applied to parents with different consequents (i.e., belonging to different subgroups). It allows migrations between niches, thus improving the search process.

The *mutation* operator randomly selects an input variable of the fuzzy rule encoded in the chromosome and 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’). Note that it is always possible to perform at least one of these options.

6. Experimentation and knowledge interpretation

6.1. Marketing model and data source used for the experimentation

Regarding other published marketing-related studies that have presented a particular artificial intelligence application in marketing,

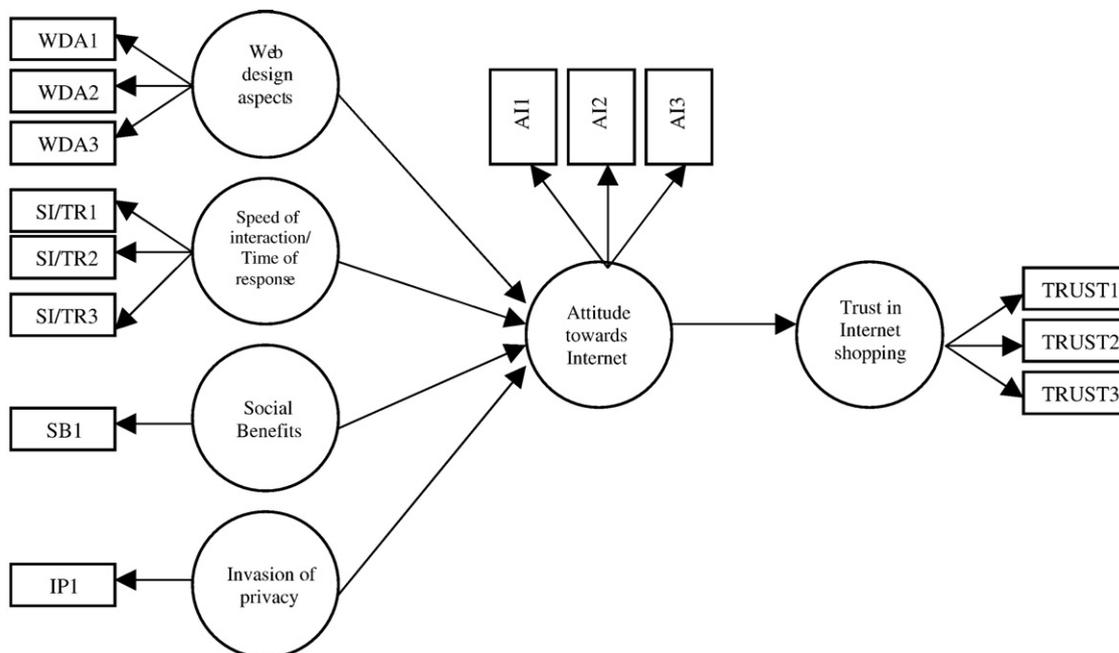


Fig. 7. Consumer behaviour model used for the experimentation (see Martínez-López & Montoro, 2005).

it is common practice to employ data and models already existing to apply to and analyse the performance of a particular KDD method or, more specifically, a data mining algorithm proposed to support a specific marketing problem (see, as e.g.: Beynon, Curry, & Morgan, 2001; Fish, Johnson, Dorsey, & Blodgett, 2004; Hurley et al., 1995; Levy & Yoon, 1995; or Rhim & Cooper, 2005).

In our case, the data employed comes from a previous research whose main findings were published in Martínez-López and Montoro (2005) and Martínez-López, Luna and Martínez (2005). The model used (see Fig. 7) was estimated using LISREL for a sample of 529 Internet users. All the variables of the model have been gathered by means of 7-point Likert-type and differential semantic scales. Taking into consideration the model of reference, we will now apply our methodology to extract descriptive fuzzy rules.

This model contains two endogenous elements/variables; i.e. attitude towards the Internet and trust in Internet shopping. Therefore, two fuzzy rule sets have to be obtained in order to explain the two endogenous concepts. The former fuzzy rule set will contain rules where the consequent is “attitude towards the Internet” and the four beliefs are considered as antecedent, while the latter will have rules with “trust in Internet shopping” as consequent and the former endogenous variable as antecedent. The fuzzy rules extracted by the proposed algorithm must be processed by the marketing expert in order to focus on the more relevant fuzzy rules to extract information about the consumer behaviour being modelled. In the next section, we illustrate how this can be done.

6.2. Post-processing: fuzzy rule selection and interpretation

6.2.1. Preliminary comments about the 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 (Fayyad & Simoudis, 1995); hence, to obtain valuable information about the problem to be solved. In this section we propose a general procedure to be applied when analysing the results (i.e. fuzzy rules) coming from the machine learning stage. In order to be concise, we next present a protocol of analysis in a structured form of four steps (subsections).

6.2.1.1. Analysis of the Pareto. The marketing expert should dedicate the first contact with the graphical representation (i.e. the full set of plots/fuzzy rules in terms of support and confidence) of the Pareto front to developing the two following preliminary tasks: (1) Analysis of the topography of every sub-Pareto front; and (2) Analysis of the composition and evolution of the absolute Pareto front.

As we have previously mentioned, the Pareto front has been defined in our methodology in function of the objectives considered by the genetic algorithm; i.e. support and confidence of the rules. The algorithm searches and generates rules in such a way that both objectives are optimised during the machine learning process. Consequently, its utility is evident as it offers information regarding the maximum frontier to achieve, hence the best/non-dominated rules, in terms of support and confidence. However, the algorithm has been designed not only to search and generate the best rules in absolute terms, but also the best rules for every category of the consequent. Hence, we distinguish between what we call the true or absolute Pareto front and the partial or sub-Pareto fronts. This method of design is much more interesting than the others that only present the best rules, as it provides information about the evolution of the rules for every category of the consequent considering their support and confidence. Moreover, to force the algorithm to explore the different subgroups helps to perform a better search process (Lavrac et al., 2004).

In sum, the graphical presentation of the Pareto front allows us to easily visualize, not only the topography of the absolute front, but also the evolution of each set of rules related to every category

of the consequent. This is of help, for instance, in seeing how the confidence evolves when the support of the rules related to a certain category of consequent increases; i.e. it allows us to visually analyse the confidence-support trade-off for every class of the consequent.

6.2.1.2. Delimitation of the rules forming the absolute Pareto front. Though considering the sub-Pareto fronts may be interesting to have a global picture of the importance of each consequent linguistic term, only those rules belonging to the absolute Pareto front (i.e., those rules whose support and confidence degrees are not simultaneously improved by any other rule) are liable to be analysed. Recall that it is not necessary for such a set of rules to share the same category of the consequent. On the contrary, it is plausible to work with an absolute Pareto front compounded by rules with diverse categories of the consequent.

6.2.1.3. Selection of the most interesting rules for the marketing expert. In general, as we have just commented, the marketing expert should focus his/her interpretative analyses for knowledge extraction on all the fuzzy rules from the absolute Pareto set. However, not all the rules of the absolute Pareto set may be statistically significant. Actually, only those rules with a high confidence degree should be considered to be reliable. Therefore, a common practice in KDD is to focus the analysis only on these rules. It is difficult to define which confidence degree is enough to be reliable, since it depends on the confidence function, the data set, and the features of the problem. Nevertheless, it is widely accepted that, in absolute terms, a rule is very reliable when its confidence is around 0.9 or higher. We will follow this procedure in the analysis performed in this paper.

Notwithstanding, we would like to point out that the subjectivity of the marketing expert can also play a significant role in the process of rules' selection (Battacharyya, 2003); for instance, determining a different confidence threshold for the selection of rules. When defining this threshold, the expert can take into account the boundaries of the absolute Pareto front (which define the highest and lowest confidence degrees that can be obtained in the analysed problem), as well as specific rules that may be of special interest for the expert. Regardless the selected threshold, the expert should have always in mind the confidence degree attached to each rule when analysing it.

Likewise, the marketing expert could also use the fuzzy rules with poor levels of confidence to rule out the information patterns shown by such rules; in some cases, it may be useful to discard a particular cause-effect relation in the variables of the model. In other words, if a certain combination of categories were found in the antecedents and consequent (i.e. causes and effect), a low level of confidence would be a clear sign of its lesser reliability.

6.2.1.4. Removing subsumed rules from the final selected fuzzy rule set. Once the final set of rules to be analysed has been delimited, it would be interesting to rationalize it if there are redundancies or the set is very large. To eliminate redundancies, it is advisable to remove those rules subsumed by others. A rule “A” is said to be subsumed by other rule “B” when, for every input variable of the antecedent, the set of linguistic terms used in the rule “A” is contained on the one used in “B”, and both rules have the same consequent. However, the marketing expert should proceed carefully, as some subsumed rule(s) may offer valuable particular information, to be also taken into account, which may be blurred in a more general rule.

6.2.2. An illustrative analysis of the fuzzy rule set on “attitude towards the Internet”

Though we are aware that a presentation of the results for every set of rules associated to our model of reference would have made this section more illustrative, we have had to rationalize it in order to not

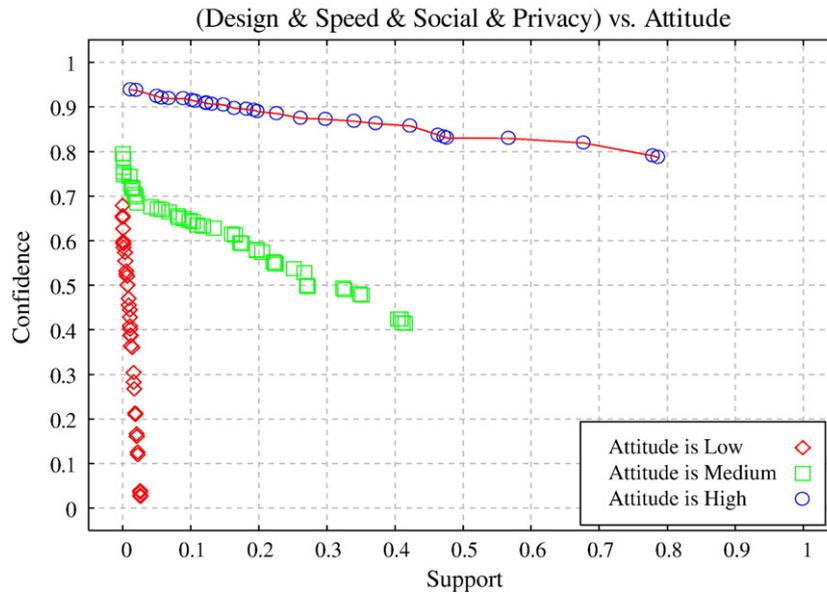


Fig. 8. Pareto front (i.e., objective values of the non-dominated solutions finally obtained by the machine learning process).

exceed the space constraints in excess. This is why we have opted to focus on just showing the reader one of the rules set generated, in particular, the one having “attitude towards the Internet” as consequent. However, this is the more complex fuzzy rule set, from the two generated, and it is very appropriate to offer the reader a good illustration of the experimental results.

6.2.2.1. Pareto front analysis. First, if we observe the Pareto front obtained for each subgroup (Fig. 8), we can see how the topography of every sub-Pareto front is quite different. There is a clear initial idea that can be extracted: the absolute Pareto front is formed by the whole set of rules whose level of attitude is high; all these rules are connected by a line. Moreover, the support-confidence trade-off is weak, inasmuch as the absolute front falls slowly from the top. Specifically, there is a loss in confidence of around 15 points from the rule with lower support up to the rule with higher; with a maximum and minimum of around 0.95 and 0.80 respectively. This means that there is a significant set of rules with considerable support, hence information patterns with wide representation in the consumers' database whose consequent is high. Hence, a straightforward conclusion is that consumers generally present good (i.e. “high”) levels of attitude towards the Internet.

With respect to the sub-Pareto fronts associated to levels of attitude low and medium, we can clearly see how their tendency is completely different. In these cases, the rules integrating such fronts rapidly lose confidence as their support increases. On the other hand, in general terms, it can be said that this question can be considered normal for those sub-Pareto fronts which are not part of the absolute Pareto fronts. For the case under analysis, this means that there are no reliable rules with levels of attitude towards the Internet low or medium. Moreover, this question is more evident as we look for rules representative of a significant mass of the consumers analysed. This fact is especially remarkable when we look at the low attitude sub-Pareto front, where there is a tremendous confidence-support unbalance.

6.2.2.2. Selection and analysis of the rules. The obtained rules are numerous. If we consider the whole set of rules for the three sub-Pareto fronts, there are over 100; i.e. the total of plots in Fig. 8. Obviously, this number strongly decreases when we consider only the constituent rules of its absolute Pareto set; i.e. 29 rules. These fuzzy rules are collected in Table 1.

To show this part of our methodology, we next analyse some example of what we consider interesting, reliable rules. However, we

insist again in the illustrative orientation of the analysis we perform. This means that a marketing expert could also consider other reliable rules (see Section 6.2.1 for more detail).

R₄: IF Design is *high* and Social is *high* and Privacy is *low* THEN Attitude is *high*; Support: 0.06, Confidence: 0.92

This rule explains the attitude towards the Internet of such consumers with a high degree of belief regarding web design aspects and social benefits, but low perception regarding Invasion of Privacy. Though the rule does not represent an important part of the population (i.e., its support is relatively low), the conclusion about these consumers having a high attitude towards the Internet is very reliable. It matches with the hypothesised relations that “design” and “social” are positively related with “attitude,” while “invasion of privacy” is negatively related. The rule could lead us to think that, considering the concurrence of the previous antecedents, the “interaction speed” is not significant to determine high levels of attitude. Furthermore, the rule R₄ is subsumed (i.e., contained) by the rule R₁₂, which will be described later in this section.

R₅: IF Speed is *medium* or *high* and Social is *high* and Privacy is *low* THEN Attitude is *high*; Support: 0.07, Confidence: 0.92

This second rule focuses on those consumers with a medium or high opinion about the interaction speed, high belief of social benefits and low perception of privacy invasion; see how in this case the antecedent speed of interaction plays an active role in determining high levels in the consequent. In particular, we can conclude with a high statistical significance that, when this scenario is observed in a consumer, he/she will have a high attitude towards the Internet. Note that, in this rule, the “design” is not decisive to infer high levels of attitude in consumers. However, it does not necessarily mean that “design” is not an influencing factor. On the contrary, if we take a look to the next rule, more general and not subsumed by any other, the information pattern it offers is quite powerful.

R₁₂: IF Design is *high* and Social is *high* THEN Attitude is *high*; Support: 0.15, Confidence: 0.90

Table 1
Pareto set of fuzzy rules that relate Design, Speed, Social and Privacy with Attitude.

	Design			Speed			Social			Privacy			Attitude			Support	Confidence
R1	-	-	High	-	-	High	-	-	High	Low	-	-	-	-	High	0.011838	0.938272
R2	-	-	High	-	-	High	-	-	High	Low	-	High	-	-	High	0.020216	0.937198
R3	-	-	High	-	Medium	High	-	-	High	Low	-	-	-	-	High	0.050510	0.923851
R4	-	-	High	-	-	-	-	-	High	Low	-	-	-	-	High	0.057738	0.919948
R5	-	-	-	-	Medium	High	-	-	High	Low	-	-	-	-	High	0.068000	0.919014
R6	-	-	High	-	Medium	High	-	-	High	Low	-	High	-	-	High	0.089232	0.918592
R7	-	-	High	-	-	-	-	-	High	Low	-	High	-	-	High	0.101966	0.914659
R8	Low	-	High	-	Medium	High	-	-	High	Low	-	High	-	-	High	0.106983	0.912913
R9	Low	-	High	-	-	-	-	-	High	Low	-	High	-	-	High	0.123097	0.909173
R10	-	-	-	-	Medium	High	-	-	High	Low	-	High	-	-	High	0.123128	0.907556
R11	-	-	High	-	Medium	High	-	-	High	-	-	-	-	-	High	0.131344	0.906223
R12	-	-	High	-	-	-	-	-	High	-	-	-	-	-	High	0.148481	0.904538
R13	Low	-	High	-	-	-	-	-	High	Low	-	High	-	-	High	0.164252	0.896915
R14	-	-	-	-	-	-	-	-	High	Low	-	High	-	-	High	0.198405	0.889739
R15	Low	-	High	-	-	-	-	-	High	-	-	-	-	-	High	0.182259	0.894958
R16	-	-	-	-	Medium	High	-	-	High	-	-	-	-	-	High	0.193198	0.892578
R17	-	-	-	-	-	-	-	-	High	-	-	-	-	-	High	0.226669	0.884859
R18	-	-	High	-	Medium	High	-	Medium	High	Low	Medium	-	-	-	High	0.261819	0.874644
R19	-	-	High	-	-	-	-	Medium	High	Low	Medium	-	-	-	High	0.298309	0.872153
R20	-	-	High	-	-	-	-	Medium	High	-	-	-	-	-	High	0.340327	0.867898
R21	Low	-	High	-	-	-	-	Medium	High	Low	Medium	-	-	-	High	0.372155	0.862637
R22	Low	-	High	-	-	-	-	Medium	High	-	-	-	-	-	High	0.422608	0.857083
R23	-	-	-	-	Medium	High	-	Medium	High	Low	Medium	-	-	-	High	0.463759	0.836779
R24	Low	-	High	-	-	-	-	-	-	-	-	-	-	-	High	0.472447	0.832564
R25	-	-	-	-	Medium	-	-	Medium	High	-	-	-	-	-	High	0.476590	0.830058
R26	-	-	-	-	-	-	-	Medium	High	Low	Medium	-	-	-	High	0.567207	0.829500
R27	-	-	-	-	-	-	-	Medium	High	-	-	-	-	-	High	0.677239	0.819155
R28	-	Medium	High	-	-	-	-	-	-	-	-	-	-	-	High	0.778952	0.790380
R29	-	-	-	-	-	-	-	-	-	-	-	-	-	-	High	0.786780	0.786780

The rule R12 is probably one of the most interesting of the set of rules with excellent levels of confidence. Its support and simplicity makes it useful for extracting conclusions about the system being modelled. This rule basically says that, regardless of the values taken from the beliefs “speed of interaction” and “invasion of privacy”, when the consumers’ opinion about the “Web design aspects” and “social benefits” are high, consumers will probably present good opinions about the Internet. So, we could conclude that, for an important portion of the population, “speed” and “privacy” are not determinant beliefs in producing high states of attitude in consumers’ minds when they have favourable opinions regarding the design aspects and social benefits of the Internet.

This is a general rule that involves several accurate rules; i.e. the first eleven rules, with the exception of rules R5, R8, R9, and R10, are subsumed in this rule. This is why we also recommend to keep the rest of reliable (i.e., with high degree of confidence) but subsumed rules of the absolute Pareto set, as they can provide particular information regarding certain influential relationship antecedent–consequent that may be omitted by a general rule.

R₁₇: IF Social is *high* THEN Attitude is *high*; Support: 0.23, Confidence: 0.88

Apart from the rules with excellent confidence described previously, the marketing expert could make use of others with slightly less reliability, though still acceptable. For instance, the rule R17—with a confidence of 0.88—shows a relation to just one belief (i.e. social benefits) and the consumers’ attitude towards the Internet. This kind of information pattern provided by our method is interesting as it is very simple and straightforward. It gives us information about the relevance of this belief, regardless of the rest, to determine high levels of attitude in the consumer. Notwithstanding, as its confidence is not within the range of excellence, this information should be used with caution.

6.3. Intrinsic and complementary benefits of our fuzzy modelling approach. Brief reflections based on the experimental results

At this stage of the paper, the whole methodology we propose has been already described and experimented. So, the reader should have an overall view of its constituent elements and performance. In particular, the previous Section 6.2.2 has aimed to illustrate the kind of information our method offers, as well as how to interpret it to extract knowledge. Nevertheless, though it has been treated in several parts of the theoretical sections, we still have not dedicated time to explicitly reflect, under the base of the empirical results of the experimentation, about the advantages of our fuzzy modelling-based method.

During the illustrative analysis of the fuzzy rules generated, the reader should have realised several of the inherent and complementary benefits provided by our method when compared with parametric statistical techniques and, in particular, with the information offered by linear regression modelling-based methods. However, we would like to dedicate the next paragraphs to briefly point out its singular advantages, with illustrative examples extracted from our experimental results.

Finally, it is important that we are aware of the fact that both modelling methods have in common giving information about how certain marketing real system behaves, making use of a model that guides the information search process in a marketing database. Notwithstanding, there are evident differences that produce a heterogeneous base of comparison between them, so absolute generalizations about the superiority of fuzzy modelling regarding with linear modelling, and vice versa, are not appropriate.

6.3.1. Linguistic instead numeric information

In order to illustrate our reflections, let us focus on the fuzzy rule R17 (included in Table 1 and previously described in Section 6.2.2.2) that relates Social Benefits and Attitude towards Internet. Which would be the information offered by linear regression modelling in this case? According

to the results reported in Martínez-López and Montoro (2005) applying SEM, the estimated coefficient (significant at $p=0.001$) for this relation is 0.36. The main information we can extract from here is, basically, that under the linear assumptions of the method, and the customer database where it has been analysed, there is a relevant and positive direct effect of the consumer's belief related to social benefits on his/her attitude towards the Internet. Also, for sure, it would not be unusual that every one of us intuitively uses an "adjective" to linguistically refer to that numeric relationship value like "strong"/"intense"/"clear", etc. However, though we perfectly understand that human beings, at last, need to make use of this "qualitative" language to process and reason, it is not correct in this case. It is easy, if you work with numbers, you cannot rigorously "translate" numerical results into words, unless you count with a sound tool (as the proposed KDD) and an appropriate language interface (as fuzzy logic). Therefore, while most people tend to mentally process the numeric information provided by the linear regression modelling in a conceptual sphere (with the subsequent context-dependent subjective interpretation), our method directly expresses this knowledge in a linguistic, but rigorous way.

Furthermore, as the reader already knows, the linguistic information provided by our method is also supported by two numeric indexes for each rule, i.e. support (degree of representation of the relation in the data set) and confidence (degree of reliability of the information pattern). The latter value, confidence, could play the role of validating the information contained in the rule. In other words, this index is useful to confirm or rule out certain scenarios of relationships between the variables of the model. In sum, these indexes allow the expert to be more accurate when making statements, under the base of fuzzy rules generated, on the marketing real system being modelled.

6.3.2. Local instead global relationships information

Following with the example of the previous section, though linear modelling says there is a coefficient of relation near to 0.4 between Social Benefits and Attitude, fuzzy rule R17 strongly suggests that such influence is localised in certain part of the range of variation of the Social Benefits (when it is "high"), where it clearly contributes to a "high" Attitude towards Internet. However, if we take a look at Table 1, there is no significant information about consumers' Attitude towards the Internet when their perception about the Social Benefits of the Internet is "low". This fact encourages the idea that the relation between two variables does not have to be the same during their whole domain.

Definitely, the real behaviour of the relation is not described under the base of an overall linear parameter of relation, as linear regression modelling assumes, since real influence may exist or be focused on a located region. Indeed, marketing academics and, especially, practitioners should not only be interested in an overall sense/direction in the relation between both variables (a common practice among the academics when formulating hypotheses, conditioned by the traditional testing methods). Moreover, it is desirable that these marketing experts could also hypothesise or speculate about the concurrence of particular values in the variables of the relationships being analysed. Our method would enable to put into practice this latter alternative.

6.3.3. Non-monotonic and multivariate relations

A researcher may have a working assumption for a specific scenario in a one-to-one relation or even better a many-to-one relationship. In fact, the relation between two variables is surely affected by other ones in the degree in which they all are correlated. Thus, the degree of the relationship between two variables can be different depending on the values of the remaining variables. Although linear regression modelling considers these correlations to estimate the relations among variables, the final provided information (i.e., estimated coefficients) summarizes this fact, thus giving a one-to-one relation degree. Besides, the relation between variables does not have to be monotonic (i.e., always increasing or decreasing), but the sign of the relation between certain variable and its cause(s) can vary depending on the values taken by its antecedent(s) variable(s). Inexplicably, the hypotheses in causal modelling traditionally do not consider this fact, although it may appear very often in the available data set. Our proposed KDD method, however, intrinsically returns non-monotonic and multivariate relations. We illustrate this feature in the following.

Let us illustrate this question with the following example. Considering the rules R9, R10 and R12 (the three are reliable and not subsumed by any other) in Table 1 we have built Fig. 9 to show an overview of the existing relations among the diverse variables considered in these rules. Since we have four input variables, it is difficult to depict the relations in two dimensions. Thus, we have elaborated a matrix that contains different combinations of two of the input variables (Social Benefits and Interaction Speed), those that seem to have a lower discrimination influence. Each cell of the matrix represents in a grey colour scale the Attitude towards Internet degree depending on the other two input variables, i.e. Invasion of Privacy and Web Design. All these values are computed by a fuzzy logic

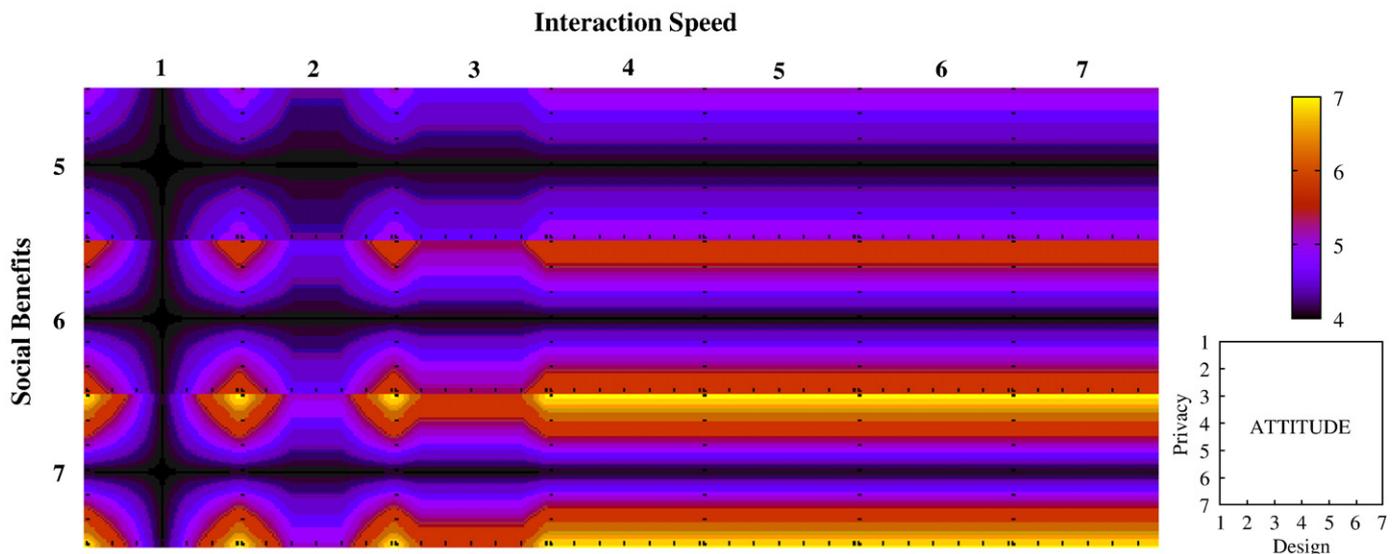


Fig. 9. Chromatic transition map generated to reflect the consumer's Attitude by the three analysed fuzzy rules.

mechanism inference for the three analysed fuzzy rules. The final result is a chromatic transition map that represents the complex existing relations at a glance.

The picture of Fig. 9 allows us to observe the non-monotonic and multivariate relations. Now, let us draw our attention on some specific examples to clarify this fact. Fig. 10 depicts the relation between Invasion of Privacy and Attitude when values for the rest of the antecedent variables are fixed. Specifically, Fig. 10(a) shows different relations depending on the value taken by the Web Design variable when Speed and Social are respectively fixed to 1 and 7. Fig. 10(b) shows the same for Speed equals to 1 and Social equals to 5; the set of values for these three variables has been fixed under the base of analysing two different scenarios of relationships between Privacy and trust, in order to better illustrate the non-monotonicity previously commented. These two plots clearly show that the degree of relationship between Privacy and Attitude varies according to the value taken by the remaining variables. Also, these figures show that this relation is not monotonic, since it takes a negative sign when Privacy is low but a positive sign when Privacy is high.

On the contrary, the information provided by the estimated coefficient obtained by SEM is simply a value, -0.15 , that summarizes the relation between Privacy and Attitude. If we graphically represent this dependency, we would obtain a function (with slope -0.15) as the one depicted in Fig. 11. The knowledge extracted from this result is by far poorer than the one depicted in Fig. 9.

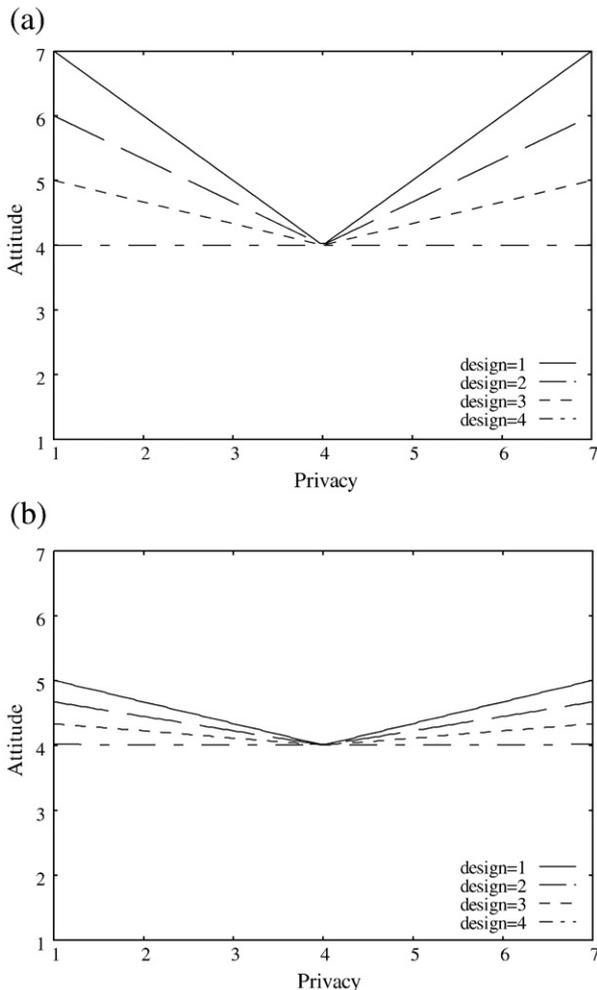


Fig. 10. Graphical representation of the relation between *Invasion of Privacy* and *Attitude towards Internet* at different cases (extracted from Fig. 9): (a) Speed=1, Social=7, and Design $\in\{1,2,3,4\}$; (b) Speed=1, Social=5, and Design $\in\{1,2,3,4\}$.

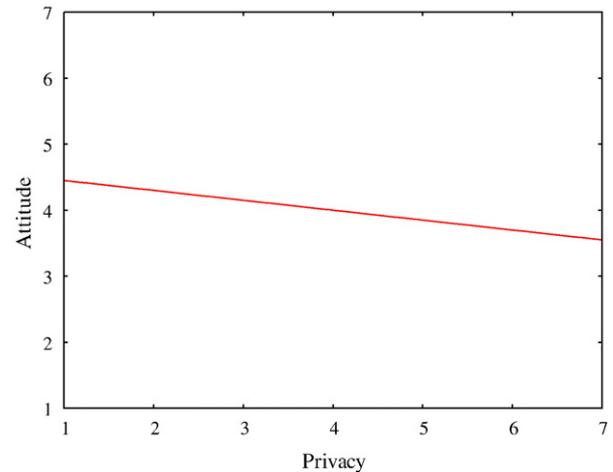


Fig. 11 Example of a graphical representation of the relation between *Invasion of Privacy* and *Attitude towards Internet* based on the standardised coefficient (-0.15 at a statistical significance $p=0.05$) obtained after applying SEM (see Martínez-López & Montoro, 2005).

6.3.4. Marketing Intelligent Systems as managerial decision support tools

Following with these reflections, we think that our method is really superior in the professional field, to be of help to support marketing decisions. Here is where parametric methods fail or, at least, do not demonstrate to perform so well. However, this is normal, considering that our method belongs to the young family of what we like to brand “Marketing Intelligent Systems,” more appropriate to support decision problems of managers.

Let us imagine that a marketing managers needs to take a decision under the base of how our customers behave taking as a base our model of reference. If s(he) had to decide between the information provided by linear regression modelling (basically, the linear coefficients) and by our method (fuzzy rules contained in Table 1), s(he) would probably feel safer with our fuzzy modelling-base method, as the perception of risk to take decisions based on such information should be lower; i.e. the information provided is richer, closer to his/her way of reasoning, s(he) can find fuzzy rules that by themselves support certain market decision or discover certain market situation not expected, etc.).

However, parametric and, in particular, linear modelling-based methods are very useful to be applied for academic purposes, especially when researchers are interested in testing theoretical causal models. This is very important for researches with a clear academic orientation; i.e. to validate proposed theoretical models. Our method cannot compete with that strength of these kinds of methods, because KDD methods were not thought and designed to test hypotheses, as statistical methods do, in the mode traditionally accepted by the scientific method. The informational philosophy of KDD methods, such as ours, is completely different. For sure, closer with what marketing managers need to support their decisions nowadays.

Notwithstanding, the academics may find other benefits, not provided by the traditional methods. We have already commented some questions above. In particular, our method gives patterns (pieces) of information which allows for better understanding the behaviour of the relation analysed, along the full range of variation. Our method does not have the inconvenient of being conditioned by a theoretical parametrical distribution which adjusts the data, neither just giving an overall coefficient of relation, as linear modelling does. This gives freedom to the data mining process, so more reliable information to the marketing analyst, as it is based on the subjacent patterns existing in the customers’ database.

All these questions are very suitable in a behavioural science as marketing, so it should be likewise of help for academic studies with a

more applied approach. Anyhow, in the academic field, we also understand the use of our method as a complement of the results obtained by using traditional modelling methods.

7. Final discussion and concluding remarks

We have faced an interesting problem with KDD in relation to marketing causal modelling and its resolution by Genetic Fuzzy Systems. The problem presents a specific type of uncertain data that justifies the use of fuzzy rules. Furthermore, we have practiced a multiobjective optimisation in order to obtain rules with high degrees of support and confidence. The KDD methodology proposed is successfully applied to a real problem of consumer behaviour modelling in online environments, where we have offered an overall perspective of how it works. The results we have obtained have been satisfactory. Summarizing, we believe that its use is very promising for academic and, especially, managerial purposes.

7.1. Theoretical contributions

This research has aimed to contribute to the marketing discipline in two differentiated but related ways. First, we have reflected on a question that may surely be controversial and give rise to certain debate, though we think that it deserves to be tackled by marketing scholars: why marketing practitioners do not make satisfactory use of the marketing models posed by the academics in their studies. It is reasonable to think that there is a gap between the concerns of academics and professionals. Anyhow, we, the academics, must invest resources to close such a gap, as the normal and successful applicability of our research should be what gives our daily work meaning. Doubtless, the right way to achieve this necessarily goes down the lines of better understanding the practitioners' needs, hence developing both theoretical models and modelling methods of analysis with a demand-side orientation. In this sense, firms are observed to have made increasing use of knowledge-driven MkMSS in the last decade to guide their market-related decision processes. Specifically, firms that have to face the analysis of huge databases, with a considerable number of variables and relations, have seen in KDD techniques and methods great potential and utility in support of their decision problems. In other words, there is a clear evolution of the recent MkMSS towards the importation or adaptation of avant-garde KDD methods from the artificial intelligence field. We like to brand such systems as "Marketing Intelligent Systems."

Hence, the academics should not only increase their efforts in importing and adapting these techniques into the marketing arena, so improving the marketing modelling methods in this line, but it would also be reasonable to expect academic studies to exploit the advantages of the artificial intelligence methods; this is a new and interesting new research stream that is likely to emerge in our discipline which deserves to be expanded. Obviously, we do not defend an indiscriminate use of these types of methods, but an intelligent use of them. This will imply, in our opinion, a synergic use of the KDD methods with the statistical approach traditionally used. For instance, in this paper (Section 6.3) we have showed how fuzzy modelling can offer the marketing expert plausible additional informational benefits, with a different approach, to the information provided by linear regression modelling.

Notwithstanding, based on our experience, we have seen certain resistance among academics in accepting its use in academic papers. Maybe there is a strong cultural barrier to surpass. Our aim, with this in mind, is not to make a subliminal suggestion that foreshadows or even apologies for a hypothetical brand new "kingdom" of artificial intelligence methods and tools, with a subsequent dethroning of the arsenal of traditional statistical techniques applied to marketing modelling and decision support. On the contrary, our approach and treatment of this question has always tried to be complementary; i.e.

every analytical method has its own characteristics for solving or supporting certain marketing decisional problem, so they may coexist. However, a definite rise in the use of these new methods in the professional arena is foreseeable, according to the current predominance of the knowledge-based MkMSS and the clear dependence of such systems on methods coming from the artificial intelligence field. Therefore, based on the necessary connection between the research interests of professionals and academics, it is not unreasonable to expect, considering such practical circumstances, an increase in proposals and use of artificial intelligence-based marketing methods by scholars. Anyhow, we suspect that there is still a long way to go. The editorial policies of marketing journals will have much to do in reversing this tendency, giving a determined chance for this new and promising research stream to be developed, discussed and matured in print.

The second and main contribution of this research is the proposal of a complete methodology to be applied in marketing causal modelling by a Genetic Fuzzy System, a specific soft computing hybridization, with a fuzzy rule descriptive induction approach. This intelligent system allows the researcher to obtain a view of the relations among variables in a new way, when compared with the kind of output we use to obtain relations from the statistical techniques in our discipline. It offers singular information patterns for every causal relation contained in the theoretical model used to guide the machine learning process. In this regard, such a process is driven by a genetic algorithm with a multiobjective optimisation approach, especially designed for proper management with the kind of measurement scales used in marketing. Furthermore, due to the benefits provided by fuzzy logic, such patterns are expressed in a way that is easily understood by all; i.e. in linguistic terms. Hence, this facilitates not only the understanding of the behaviour of the relations among the variables of the real marketing system being modelled, but it also allows us to find interesting scenarios in the database analysed which would not be possible to see with parametric estimation techniques.

Finally, we would like to point out the following challenges we have had to face during the development of the methodology:

- We have reflected on how to process the kind of data that marketing researchers and professionals usually work with; i.e. several indicators or items related to a certain element (unobserved variable) of the marketing model. In this sense, instead of what could first be done intuitively, i.e. treating them at a pre-processing stage of the KDD process, we have proposed an original and reasonable solution that allows the treatment of the data during the machine learning process. We have called this procedure "multi-item fuzzification."
- We have reflected on and proposed reasoned solutions to the problem of transformation of the marketing scales into linguistic variables. To the best of our knowledge, this is also an original contribution of our research.
- A genetic algorithm has been designed *ad hoc* for the marketing problem we have faced in this research. In this regard, we discuss the right optimisation approach to follow.
- As far as we know, there are no previous studies that have proposed a methodology similar to ours, so we have had to think about how to analyse, interpret and extract knowledge from the results offered by the machine learning process. To do that, we have proposed a reasoned and detailed protocol of analysis.

7.2. Managerial implications

The incremental benefits that this method offers to the managers could be synthesised in the following:

- Managers have a powerful method for customer database analyses which, based on a causal model that guides the searching process, mixes the suitability of fuzzy rules in order to process the kind of original data marketers usually work with and to express it with

subjective concepts, emulating human reasoning. This is achieved through the effectiveness and accuracy of genetic algorithms to find optimum solutions, in this case, individual information patterns about the customers' database, expressed in qualitative/linguistic terms. Consequently, the managers can support their decisions with information output expressed in a way similar to how they think and talk. This kind of information is highly appropriate for supporting non-structured problems the marketing managers may have to face.

- Such linguistic variables have a basis in the Fuzzy Set Theory, so the managers can associate, based on their own experience and criteria, subjective concepts—such as: low, high, loyal, risky, etc—to a particular range of quantitative values defining the set of variables the database contains.
- This method allows the simultaneous working with marketing variables originally gathered through different measurement scales; i.e. either nominal, ordinal, interval or ratio scales.
- Thanks to the self-defining characteristics of the genetic algorithms, it is a robust and reliable method when applied to large databases; i.e. it is scalable.
- Last, but not least, this methodology has undergone experimentation with a consumer model, so that it can be applied in supporting managers' decisions on consumers/customers. However, it would be equally valid to support other facets of the marketing practitioners' sphere of decisions. That all depends on the focus of the causal model—i.e. the antecedents and the consequents it contains—and the database we use as reference. For instance, we could apply this method, thus taking advantage of its characteristics, to support decisions regarding product design, advertising, prices policies, etc.

8. Closing comments: limitations and future research opportunities

Among the main limitations we identify here is the real application of this methodology we present in this paper. We would understand those readers, either scholars or practitioners, who after assimilating the method we propose here said something like: “right, but how could I apply this tomorrow to my research (scholar), or to my decisional process (practitioner)?” Obviously, it would be necessary to have specific software. Such software has already been designed and developed by us; we would not have been able to empirically test our method without it. Notwithstanding, it is not developed enough yet for commercialization.

In sum, these are some of the questions that encourage us to go on with the main research project underlying this paper. In particular, some of the research opportunities, hence new challenges to tackle, that are occupying our time in the near future are the following:

- Improvement of the genetic algorithm used in the machine learning stage, in order to further improve the performance and accuracy of the fuzzy rules discovery process.
- Design and application of new metrics/indexes, added to those of support and confidence, to better evaluate the fuzzy rules obtained; for instance, metrics related to the interestingness of the rules.
- The current method has been designed to drive the machine learning process by using a marketing causal model to interconnect variables in the space of search; i.e. what is called “supervised learning”. However, sometimes, the manager or the academic may not have full information about any relation structure among the variables containing a particular database. In other words, the marketing expert may know certain relations, though (s)he may not be aware of reasonable relations about others. Moreover, there could not be aprioristic information about the relations of the variables, or even an attempt to search, without any restriction of search imposed by any model, for “covered” structures in the database. In this case, we could develop what is known as semi-supervised and unsupervised learning, respectively.
- Finally, we are working on designing user-friendly software to apply this method. Specifically, this software is integrated in a wider

research project focused on developing a software package of diverse artificial intelligence tools to be applied in KDD, called KEEL (<http://www.keel.es>). This project is supported by the Spanish Ministry of Education and Science.

Acknowledgements

The authors wish to deeply thank the Editor-in-Chief, Peter J. Laplaca, and the anonymous reviewers for their high valuable comments and orientations. This work has been supported in part by the Spanish Ministry of Science and Education under grant no. TIN2005-08386-C05-01 and by the Andalusian Government under grant no. P07-TIC-3185.

References

- Agarwal, A., Shankar, R., & Tiwari, M. K. (2007). Modelling agility of supply chain. *Industrial Marketing Management*, 36(4), 443–457.
- Battacharyya, S. (2003). Evolutionary computation for database marketing. *Journal of Database Management*, 10(4), 343–352.
- Beynon, M., Curry, B., & Morgan, P. (2001). Knowledge discovery in marketing. An approach through rough set theory. *European Journal of Marketing*, 35(7/8), 915–935.
- Bucklin, R. E., Lehmann, D. R., & Little, J. D. C. (1998). From decision support to decision automation: A 2020 vision. *Marketing Letters*, 9, 234–246.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: from concept to implementation*. Prentice Hall.
- Campbell, A. J. (2003). Creating customer knowledge competence: managing customer relationship management programs strategically. *Industrial Marketing Management*, 32(5), 375–383.
- Carlsson, C., & Turban, E. (2002). Introduction. DSS: Directions for the next decade. *Decision Support Systems*, 33, 105–110.
- Casillas, J., Cordon, O., Herrera, F., & Magdalena, L. (Eds.). (2003a). *Interpretability issues in fuzzy modeling*. Heidelberg, Germany: Springer.
- Casillas, J., Cordon, O., Herrera, F., & Magdalena, L. (Eds.). (2003b). *Accuracy improvements in linguistic fuzzy modeling*. Heidelberg, Germany: Springer.
- Casillas, J., Martínez-López, F. J., & Martínez, F. J. (2004). Fuzzy association rules for estimating consumer behaviour models and its application to explain trust in Internet shopping. *Fuzzy Economic Review*, IX(2), 3–26.
- Chankong, V., & Haimes, Y. Y. (1983). *Multiobjective decision making theory and methodology*. North-Holland.
- Christopher, M. (2000). The agile supply chain: Competing in volatile markets. *Industrial Marketing Management*, 29(1), 37–44.
- Coello, C., Van Veldhuizen, D. A., & Lamont, G. B. (2002). *Evolutionary algorithms for solving multiobjective problems*. New York: Kluwer Academic Publishers.
- D'Aveni, R. A. (1994). *Hypercompetition – Managing the dynamics of strategic maneuvering*. New York, NY: The Free Press.
- Deb, K., Pratap, A., Agarwal, S., & Meyarevian, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.
- Drejer, A. (2004). Back to basics and beyond. Strategic management – An area where practice and theory are poorly related. *Management Decision*, 42(3/4), 508–520.
- Dubois, D., Prade, H., & Sudkamp, T. (2005). On the representation, measurement, and discovery of fuzzy associations. *IEEE Transactions on Fuzzy Systems*, 13(2), 250–262.
- Eliashberg, J., & Lilien, G. L. (1993). Mathematical marketing models: Some historical perspectives and future projections. In J. Eliashberg & G.L. Lilien (Eds.), *Handbooks in operations research and management science/Marketing*, Vol. 5. (pp. 3–23). Amsterdam: North-Holland.
- Fayyad, U. M., Piatesky-Shapiro, G., Smyth, S., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. M.I.T. Press.
- Fayyad, U. M., & Simoudis, E. (1995). *Knowledge discovery and data mining: Tutorial. 14th International Joint Conference on Artificial Intelligence*. Canada: Montreal.
- Fish, K. E., Johnson, J. D., Dorsey, R. E., & Blodgett, J. G. (2004). Using an artificial neural network trained with a genetic algorithm to model brand share. *Journal of Business Research*, 57(1), 79–85.
- Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. Heidelberg, Germany: Springer.
- Gatignon, H. (2000). Commentary on Peter Leeflang and Dick Wittink's “Building models form marketing decisions: Past, present and future”. *International Journal of Research in Marketing*, 17, 209–214 (Special issue on marketing modelling on the threshold of the 21st century).
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA, USA: Addison-Wesley.
- González, A., & Pérez, R. (1998). Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1), 37–51.
- Han, J., & Kamber, M. (2001). *Data mining. Concepts and techniques*. Morgan Kaufmann Publishers.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Michigan, MI, USA: Ann Arbor: The University of Michigan Press.
- Hurley, S., Moutinho, L., & Stephens, N. M. (1995). Solving marketing optimization problems using genetic algorithms. *European Journal of Marketing*, 29(4), 39–56.

- Lavrac, N., Cestnik, B., Gamberger, D., & Flach, P. (2004). Decision support through subgroup discovery: Three case studies and the lessons learned. *Machine Learning*, 57(1–2), 115–143.
- Leefflang, P. S. H., & Wittink, D. R. (2000). Building models for marketing decisions: past, present and future. *International Journal of Research in Marketing*, 17, 105–126 (Special issue on marketing modelling on the threshold of the 21st century).
- Leefflang, P. S. H., Wittink, D. R., Wedel, M., & Naert, P. A. (2000). *Building models for marketing decisions*. Kluwer Academic Publishers.
- Levy, J. B., & Yoon, E. (1995). Modeling global market entry decision by fuzzy logic with an application to country risk assessment. *European Journal of Operational Research*, 82, 53–78.
- Li, S., Davies, B., Edwards, J., Kinman, R., & Duan, Y. (2002). Integrating group Delphi, fuzzy logic and expert systems for marketing strategy development: the hybridization and its effectiveness. *Marketing Intelligence & Planning*, 20(5), 273–284.
- Li, S., Kinman, R., Duan, Y., & Edwards, J. S. (2000). Computer-based support for marketing strategy development. *European Journal of Marketing*, 34(5/6), 551–575.
- Lilien, G. L., Kotler, P., & Moorthy, K. S. (1992). *Marketing models*. Prentice-Hall International Editions.
- Lin, Y., Su, H. Y., & Chien, S. (2006). A knowledge-enabled procedure for customer relationship management. *Industrial Marketing Management*, 35(4), 446–456.
- Lindskog, P. (1997). Fuzzy identification from a grey box modeling point of view. In Hellendoorn, & Driankov (Eds.), *Fuzzy model identification* (pp. 3–50). Heidelberg, Germany: Springer-Verlag.
- Little, J. D. C. (1970). Models and managers: The concept of a decision calculus. *Management Science*, 16(8), B466–B485.
- MacLean, S., & Gray, K. (1998). Structural equation modelling in market research. *Journal of the Australian Market Research Society*, 6, 17–32.
- Martínez-López, F. J., & Casillas, J. (2007). Consumer modeling by multiobjective Genetic Fuzzy Systems: A descriptive rule induction approach. *Proceedings of the 36th European Marketing Academy (EMAC) Conference, Reykjavik (Iceland)*.
- Martínez-López, F. J., Luna, P., & Martínez, F. J. (2005). Online shopping, the standard learning hierarchy, and consumers' Internet expertise. An American–Spanish comparison. *Internet Research*, 15(3), 312–334.
- Martínez-López, F. J., & Montoro, F. J. (2005). Modelling consumer trust in internet shopping based on the standard learning hierarchy: A structural Approach. *Journal of Internet Business*, 2 (April). Available from http://www.bizfac.com/jib/iss02_lopez.pdf
- Metaxiotis, K., Psarras, J. E., & Samouilidis, J. E. (2004). New applications of fuzzy logic in decision support systems. *International Journal of Management and Decision Making*, 5(1), 47–58.
- Michalewicz, Z. (1996). *Genetic algorithms + data structures = evolution programs*, 3rd ed. Heidelberg, Germany: Springer-Verlag.
- Mitra, S. (2002). Data mining in soft computing framework: A survey. *IEEE Transactions on Neural Networks*, 13(1), 3–14.
- Nissen, V. (1995). An overview of evolutionary algorithms in management applications. In J. Biethahn & V. Nissen (Eds.), *Evolutionary algorithms in management applications* (pp. 44–100). Springer-Verlag.
- Rhim, H., & Cooper, L. G. (2005). Assessing potential threats to incumbent brands: New product positioning under price competition in a multisegmented market. *International Journal of Research in Marketing*, 22, 159–182.
- Roberts, J. H. (2000). The intersection modelling potential and practice. *International Journal of Research in Marketing*, 17, 127–134.
- Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present and future of decision support technology. *Decision Support Systems*, 33, 111–126.
- Steenkamp, J., & Baumgartner, H. (2000). On the use of structural equation models for marketing modelling. *International Journal of Research in Marketing*, 17, 195–202 (Special issue on marketing modelling on the threshold of the 21st century).
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Stevens, S. S. (1959). Measurement, psychophysics and utility. In C. W. Churchman, & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 18–63). New York: John Wiley.
- Talvinen, J. M. (1995). Information systems in marketing: Identifying opportunities for new applications. *European Journal of Marketing*, 29(1), 8–26.
- Van Bruggen, G. H., & Wierenga, B. (2000). Broadening the perspective on marketing decision models. *International Journal of Research in Marketing*, 17, 159–168 (Special issue on marketing modelling on the threshold of the 21st century).
- Viswanathan, M., Bergen, M., Dutta, S., & Childers, T. (1996). Does a single response category in a scale completely capture a response? *Psychology & Marketing*, 13(5), 457–479.
- Wedel, M., Kamakura, W., & Böckenholt, U. (2000). Marketing data, models and decisions. *International Journal of Research in Marketing*, 17, 203–208 (Special issue on marketing modelling on the threshold of the 21st century).
- Wierenga, B., & Van Bruggen, G. T. (1997). The integration of marketing problem-solving modes and marketing management support systems. *Marketing Management Support Systems*, 61(July), 21–37.
- Wierenga, B., & Van Bruggen, G. T. (2000). *Marketing management support systems: Principles, tools and implementation*. Kluwer Academic Publishers.
- Wind, Y., & Lilien, G. L. (1993). Marketing strategy models. In J. Eliashberg, & G. L. Lilien (Eds.), *Handbooks in operations research and management science: Marketing* (pp. 773–826). Amsterdam (The Netherlands): Elsevier Science Publishers.
- Winer, R. S. (2000). Comment on Leefflang and Wittink. *International Journal of Research in Marketing*, 17(Special issue on marketing modelling on the threshold of the 21st century), 141–146.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8, 338–353.
- Zadeh, L. A. (1975a). The concept of linguistic variable and its application to approximate reasoning (Part I). *Information Sciences*, 8, 199–249.
- Zadeh, L. A. (1975b). The concept of linguistic variable and its application to approximate reasoning (Part II). *Information Sciences*, 8, 301–357.
- Zadeh, L. A. (1976). The concept of linguistic variable and its application to approximate reasoning (Part III). *Information Sciences*, 9, 43–80.

Francisco J. Martínez-López, MSc in Marketing, European PhD in Business Administration, is an Associate Professor in Marketing at the University of Granada (Spain) and Assistant Professor in Marketing at the Open University of Catalonia (Spain). He has been a visiting researcher at the Marketing Departments of the Aston Business School (Birmingham, UK) and the Michael Smurfit School of Business (Dublin, Ireland). Among his main areas of interest are consumer behaviour on the Internet, e-marketing, marketing channels and KDD methodologies for marketing. He has authored several books, book chapters, international journal and conference papers. He has also edited an international journal special issue.

Jorge Casillas received the MSc and PhD graduate degrees in Computer Science in 1998 and 2001, respectively, from the University of Granada, Spain. He is an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada. He has edited two international books, edited two international journal special issues, and organized four special sessions in international conferences. He is an author of more than 20 journal papers, 10 book chapters, and 40 conference papers. He serves on the Editorial Board of the *Evolutionary Intelligence* journal of Springer from January, 2008. He is the treasurer of the European Society for Fuzzy Logic and Technologies (EUSFLAT) and coordinator of the working group on Genetic Fuzzy Systems from September, 2007. His research interests include fuzzy modelling, intelligent robotics, Marketing Intelligent Systems, knowledge discovery, and metaheuristics.