A knowledge discovery method based on genetic-fuzzy systems for obtaining consumer behaviour patterns. An empirical application to a Web-based trust model

Jorge Casillas

Department of Computer Science and Artificial Intelligence,
Computer and Telecommunication Engineering School,
University of Granada,
Granada, E-18.071, Spain
E-mail: casillas@decsai.ugr.es

Francisco J. Martínez-López*

Department of Marketing,
University of Granada,
Granada, E-18.071, Spain
E-mail: fjmlopez@ugr.es
*Corresponding author

Abstract: This paper shows part of a larger interdisciplinary research focused on developing artificial intelligence-based analytical tools to aid the marketing managers' decisions on consumer markets. In particular, here it is presented and tested a knowledge discovery methodology based on genetic-fuzzy systems – a Soft Computing (SC) method that jointly makes use of fuzzy logic and genetic algorithms – to be applied in marketing modelling. Its characteristics are very coherent with the requirements that marketing managers currently demand to market analytical methods. Specifically, it has been paid attention to illustrate, in detail, how this proposed (Knowledge Discovery in Databases) KDD method performs with an empirical application to a Web-based trust consumer model.

Keywords: consumer modelling; web-based trust; knowledge discovery; fuzzy logic; genetic algorithms.

Reference to this paper should be made as follows: Casillas, J. and Martínez-López, F.J. (2009) 'A knowledge discovery method based on genetic-fuzzy systems for obtaining consumer behaviour patterns. An empirical application to a web-based trust model', *Int. J. Management and Decision Making*, Vol. 10, Nos. 5/6, pp.402–428.

Biographical notes: 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 been Visiting Research Fellow at the University of the West of England (Bristol, UK). He has edited two international books, edited two international journal special issues, and organised four special sessions in international conferences on the topics "interpretability-accuracy trade-off in fuzzy modeling", "genetic fuzzy systems", and "intelligent robotics". He is author of about 20 journal papers, ten book chapters, and 40 conference papers.

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 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.

1 Introduction

1.1 Background

In the past decades 'information' has been increasingly considered a strategic resource whose adequate management has arisen as a key factor to explain entrepreneurial success (see, as e.g., Ein-Dor and Jones, 1985; Evans and Wurster, 1997; Laudon and Laudon, 1998). The quantity and quality of relevant information that firms own when facing their decision-making processes has a fundamental role in ameliorating the degrees of uncertainty usually characterising the business environment. Moreover, if we attend to the considerable turbulence and competitivity defining the sectors in which companies operate, it is also necessary to have timely information, obtained as a result of an efficient process (Hannula and Pirttimaki, 2003). The accomplishment of all these questions, therefore, contributes to a reinforcement of both the firms' dynamicity and competitive response; of course, along with other factors related to the philosophy of marketing strategies, planning and execution that are beyond the research purposes we tackle here. In essence, it would not be difficult to achieve a general agreement with the asseveration that the above ideas are the root cause of the appearance, improvement and evolution of Management Information Systems (MIS) in general, and Marketing Management Support Systems (MkMSS) in particular.

How to improve the impact and productivity of marketing actions addressed to customers has been and is especially now, or at least it should be, a constant concern of firms. This question explains, for instance, the transformation undergone by the philosophy of the marketing function in the early 1980s, when the selling concept subjacent to the selling orientation evolved into the marketing concept characterising the notorious market orientation. Doubtless, the market orientation has been the predominant business philosophy from that time until now (Heiens, 2000). A successful market orientation necessarily requires that a firm worries about gathering and disseminating good and time-relevant market information to the different constituent parts of the organisation (Hunt and Morgan, 1995; Kohli and Jaworski, 1990; Narver and Slater, 1990).

Basically, firms looking to be market oriented ought to have information about their customers, competitors and markets, and make intelligent use of it when addressing their actions (Cravens et al., 2002). In this sense, as the amount of data to be accessed, stored and processed is usually so vast, firms necessarily have to make harmonised use of information technologies and MkMSS to be competitive in the current information-intensive marketing contexts (Berthon and John, 2006). Moreover, inasmuch as it is expected that the market orientation philosophy evolves towards a more specific customer-centric way of managing the exchanges with markets (Kumar and Ramani, 2006), working with precise information and decision-support systems will become even more indispensable for the firm's survival. Specifically, this changing tendency should be better seen in the Web-based market contexts, where the medium's technology facilitates individualised interaction with customers (Hoffman and Novak, 1996), as well as access and facilities to obtain consumers' data (Lan, 2006; Smith, 2005). Consequently, the Web emerges as a very appealing data source for firms which, thanks to appropriate use of information technologies, may improve their knowledge about customers, thus fostering more value-added marketing actions and offers (see, as e.g., Aziz and Yasin, 2004; Wood, 2001).

In any case, regardless of the firms' market context, either the marketplace, the market space or even both simultaneously, it seems evident that the main problem currently faced by market-oriented organisations is not the availability of information (data), but the possession of the necessary level of knowledge to take adequate decisions. Moreover, not only the possession but also the use that firms make of such knowledge is said to be a notable factor in determining their competitive position (Silvi and Cuganesan, 2005). Currently, there is no doubt about the impact that a suitable management of knowledge must have on the firms' performance. In fact, Vargo and Lusch (2004, p.9), when proposing the set of foundational premises defining the new dominant logic of marketing, highlight knowledge as the underlying principle of "competitive advantage and economic growth and the key source of wealth".

This question explains why the analytical capabilities integrating the firm's MkMSS are so crucial for businesses in this moment of time. It is likely to expect that their variety and capabilities should have a relevant role in supporting the marketing managers' decisional processes, as they should allow the achievement of high value-added information about the customers' behaviour. In this regard, it is been observed how the so-called knowledge-based MkMSS have increasingly taken a predominant position in the diversity of systems used by managers when supporting their decisions, from the middle 1980s to the present time (Shim et al., 2002; Wedel et al., 2000). These are usually based on methods for knowledge extraction from databases, which come from the computer science and artificial intelligence discipline; i.e., marketing systems based on, as e.g., artificial neural networks, genetic algorithms, fuzzy rules-based methods, etc. (Li et al., 2000; Wierenga and van Bruggen, 2000). Likewise, it is plausible to see an evolution, in the near future, towards the use of knowledge-based systems based on hybridisations – i.e., Soft Computing (SC) methodologies – of those artificial intelligence methods (Carlsson and Turban, 2002).

In sum, considering the previous paragraph as a backdrop, this paper introduces and empirically shows a knowledge discovery methodology to extract useful information patterns from consumers' databases, with a descriptive rule induction approach based on certain SC method (genetic-fuzzy systems). In this regard, we offer specific solutions to the diversity of questions related to such a Knowledge Discovery in Databases KDD process (e.g., how to process the original data, the design of the data mining state, the post-processing of the outputs, etc.). Hence, this is not a mere application of a data mining paradigm to a database. On the contrary, this research represents an important effort to adapt KDD solutions to the problem we face.

Specifically, we make use of a supervised learning algorithm in the data mining process, as we work in the scenario that the marketing manager counts on the aprioristic information of a theoretical consumer behaviour model. Basically, this drives the search process of the algorithm in the database. The practical experimentation of the methodology we propose is made by its application to an existing Web-trust consumer model, using a real database of internet users, which was originally estimated using Structural Equation Modelling (SEM). This initially allows us to show its potential to offer relevant pieces of information about the consumers being analysed, in a linguistic format which is quite understandable and helpful for supporting the managers' decisions. In the next section, we describe the concrete theme with more detail as well as the main research objectives of the paper.

1.2 Research scope and main goals

Marketing professionals and academics have usually relied on both models and the diversity of methods integrating multivariable statistical techniques to drive analyses and try to obtain information from the consumers' data. In this regard, models of consumer behaviour have demonstrated that they are, over time, not only a basis for the development of the marketing science (van Bruggen and Wierenga, 2000), but also, from a practical perspective, a useful tool to support the resolution of complex and ill-structured marketing problems (Talvinen, 1995). However, though the classic statistical techniques are still intensively used in academic researchers, the practitioners have shown a clear tendency towards the use of more powerful tools of analysis in the last two decades, based on artificial intelligence methods, capable of adequately managing the extraction of knowledge in current databases (Wierenga and van Bruggen, 2000). Thus, considering the necessary connection between the academic and professional arena, it is not unexpected that marketing academics increasingly make a greater use of these methods in the near future. With respect to this prediction, Roberts (2000), when reflecting on the present and future of marketing modelling, emphasises the necessary improvement of the techniques and methods of estimation traditionally used for academic purposes. It seems clear that the classic statistical techniques integrating the academics' arsenal of analytical capabilities needs to be refreshed in this way. With this aim, it is expected that the MkMSS trend will improve their performance by taking advantage of the synergies produced by the integration of modelling estimation techniques based on classic econometric with new methods and systems based on artificial intelligence (Gatignon, 2000; van Bruggen and Wierenga, 2000). This is the general framework where we develop our research.

The analytical capabilities to extract information from a database are multiple and with diverse properties, either by referring to the statistical or the advanced and modern data mining tools. Every tool has its own characteristics, which make it adequate to tackle certain scenarios of estimation (Hair et al., 1998); i.e., number of variables, type of relations among dependent and independent variables and measurement scales used for each of them. This obviously determines the information problem they can support. In particular, our interest research focuses on KDD guided by complex marketing models. These are causal models frequently used by consumer behaviour modellers, with several independent and dependent variables, which present a diversity of relations among them. Such a type of models has been traditionally estimated by SEM; i.e., the only statistical tool for multivariate analyses which can parameterises these models. With respect to the set of statistical tools, SEM is considered to be one of the most widely known advances used in social sciences and, specifically, in the marketing field, to test causal models (Baumgartner and Homburg, 1996). Thanks to its characteristics, using SEM, modellers can obtain diverse benefits, as e.g., to simultaneously consider several endogenous variables, to work with latent construct (also called unobserved variables) which are inferred by means of indicator/observed variables, to provide measurement errors for constructs, etc. (Bollen and Long, 1993).

Nevertheless, though SEM has been frequently applied in academic studies, to test the sense of the hypothesised relations in a theoretical model, marketing practitioners currently make less use of it (Martínez-López et al., 2006). Specifically, SEM has some shortcomings which make it a poor method to be applied by managers when trying to find valuable information to support their decisions on markets. Next, we make some commentaries with regard to their relevant weaknesses. SEM generally supposes linear relationships among the constituent elements of the model (Mueller, 1996). This is a characteristic that seriously limits its utility as a decisional support statistical tool. As we have already noted, the estimation techniques based on SEM are useful to help the academics in the task of validating the theoretical models they propose. Specifically, the main capability of this method, thus its main attraction, is to test causal relationships among the diversity of constructs (with multiple measurement items) integrating a model (Jöreskog and Sörbom, 1993). However, to put simply, it does not allow the analysing or, therefore, the interpreting of relationships among several variables when such relationships are contemplated with different degrees of intensity. Besides, though there is a recent interest in extending the SEM procedures to allow interactions and non-linearities among the variables of the models, with new variants in the methodological issues of SEM, its application and interpretation of the results are tedious, problematic, as well as being an error-prone process (Laplante et al., 1998; Tate, 1998). Anyhow, even considering these recent advances, we would still have to work with the information provided by overall 'structural' parameters of relations among variables. Thus, even if some of them were non-linear (e.g., quadratic), the main inconvenience commented on in this paragraph would still remain. In essence, all these questions make SEM a method of poor utility when used as a decision-support tool (Laurent, 2000; Steenkamp and Baumgartner, 2000).

Academics have suggested refreshing the current statistical methods of estimation applied to complex consumer behaviour models by also focusing on artificial intelligence and KDD methodologies (Wedel et al., 2000). This step would allow the improvement of the marketing modelling methods of estimation, while making more attractive their use for practitioners. Specifically, it has been underlined the potential that fuzzy modelling methods may have in the enrichment of the decision-support systems in general (Metaxiotis et al., 2004) and the MkMSS' analytical tools in particular (Li et al., 2000). These methods present several characteristics which make convenient their use in KDD processes whose data mining stage is supervised by a marketing model of reference. Thus, fuzzy rules-based KDD methods are flexible, interactive and more oriented to offer qualitative information than other statistical estimation methods used in marketing modelling (Gatignon, 2000; van Bruggen and Wierenga, 2000).

Regarding the previous introductory ideas, we would like to conclude this section by briefly describing the main challenges we have pursued with this paper:

- Reflecting on the necessary evolution of the MkMSS, towards the importation of tools and methods from the artificial intelligence field, more adequate to support the managers' decisions.
- Researching in a new KDD methodology based on fuzzy association rules to be applied in consumer behaviour modelling, able to provide useful information and good support to the marketing managers' decision-making. In this regard, we are interested in methods which are robust, reliable and able to work with big customers' databases, if necessary.
- Experimenting with the methodology we propose, to empirically show how it performs. Specifically, we use a trust-Web model application with a real database on online consumers.
- Of secondary importance, though still interesting, we also want to offer a view of how our method complements the results obtained by applying SEM.

We address these questions as follows. First, as we have just introduced, we reflect 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 (see the preceding paragraphs). In Section 2, we show how KDD methods based on fuzzy rules are a very interesting route to improve the understanding of the consumer and, specifically, to be applied in consumer behaviour modelling. Section 3 briefly presents a brand new KDD methodology for marketing modelling by genetic-fuzzy systems. Next, in Section 4 we empirically show some illustrative results. Finally, we discuss the main contributions of our research, reflecting on the academic and managerial implications.

2 Knowledge discovery based on fuzzy rules

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 et al., 1996). Furthermore, KDD implies the development of a process which is made up of several stages. In this sense, data mining, which is considered as the core of KDD process, is characterised by the application of machine learning methods to automatically or semi-automatically extracted patterns or models from data (Witten and Frank, 1999).

Nowadays, one of the most successful tools to develop 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 with the following structure:

IF X_1 is A_1 and … and X_n is A_n THEN Y_1 is B_1 and … Y_m is B_m.

These rules set logical relationships among variables of a system by using qualitative values. Such a representation mode can be easily understood by human beings. Hence, the performance of both, analysis and interpretation steps of the modelling process improve; thanks to the true behaviour of a system that is more effectively revealed. Notwithstanding, it should be noted that though human reasoning may manage without strain terms like *high* or *will rise quickly*, when this issue is tackled by means of an automatic process its treatment is more complex.

To properly work with these kinds of qualitative valuations, linguistic variables (Zadeh, 1975) based on both Fuzzy Sets Theory and Fuzzy Logic (Zadeh, 1965) are used, thus the previous 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. As an example, we consider the linguistic variable *age*, which could take the linguistic values *teenager*, *young*, *adult* and *old*.

Fuzzy rules can be employed as a kind of knowledge representation when discovering 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 examined data. In knowledge discovery, the process to obtain these patterns must be automatic, or semi-automatic, discovered patterns must be comprehensible and they must provide useful information, and data must be invariably presented in substantial quantities (Witten and Frank, 2000).

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 generated patterns are represented by an easily examined structure, 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 in explaining something about the data. This problem with KDD, the interpretability-accuracy trade-off, is also being currently faced in fuzzy modelling (Casillas et al., 2003a, 2003b) and will be considered by our proposal.

The use of fuzzy rules when developing the knowledge discovery process has some advantages, as follows: they allow us to use uncertainty data; they adequately consider multivariable relationships; results are easily understandable by a human being; additional information can be easily added by an expert; the accuracy degrees can be easily adapted to the problem necessity; the process can be highly automatic with low human intervention.

Therefore, we use fuzzy logic as a tool to structure the information of a consumer behaviour model in a clear, legible way that is close to the human being. The fuzzy system allows us to properly represent the interdependence of variables and the non-linear relationships that could exist among them. Finally, optimisation algorithms (a genetic algorithm in this paper) are used to design the fuzzy rules to meet the interpretability and accuracy criteria imposed by the expert. The following section introduces the methodology used for discovering knowledge by means of fuzzy rules and genetic algorithms to consumer behaviour modelling.

3 A knowledge discovery method for consumer behaviour modelling with fuzzy rules

This section introduces the process we propose to perform the KDD method applied to consumer behaviour modelling by fuzzy rules. Synthetically, it is based on preparing the data and on fixing the scheme we follow to represent the knowledge existing in the data. Once these aspects are defined, a machine learning method can be designed and used to automatically extract interesting fuzzy rules. Finally, an interpretation of the FRS generated after the data mining process is necessary to achieve high-valued information.

3.1 Data gathering

The first step is to collect the data related to the variables defining the theoretical model of consumer behaviour proposed. In this sense, as market modelling has been traditionally done, data are obtained by means of a questionnaire in a similar way to the models estimated by SEM.

3.2 Data processing

Secondly, 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 consumer behaviour modellers face and develop the measurement process of variables that complex behavioural models contain. In this respect, reflections about the measurement of such variables, with a special focus on those usually known as theoretical constructs, should be made. Consequently, we think that time should be spent on analysing the adaptation of the fuzzy rules-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 could be classified in two groups depending on if they declared that these constructs could or could not be perfectly measured by means of observed variables (indicators) – i.e., the existence or not of a one-to-one correspondence between a construct and its measurement. Certainly, though consumer behaviour modellers and marketing modellers in general, tended to make use in the beginning of what was known as the operational definition philosophy, a more convenient and reasonable position is the one more recently based on the partial interpretation philosophy which distinguished between unobserved (constructs) and observed (indicators) variables. This latter approach of measurement, being currently predominant in the marketing modelling discipline, recognises the impossibility of doing perfect measurements of theoretical constructs by means of indicators, so it poses to jointly consider multiple indicators – imperfect when considered individually, though reliable when considered together – of the subjacent construct to obtain valid measures (Steenkamp and Baumgartner, 2000).

Therefore, our methodological approach should be aware of this question when adapting the data (observed variables) to a fuzzy rule learning method. In this regard, 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 variable or an indicator associated with each of them, even when they have been measured with diverse measurement scales. The problem comes, hence the challenge, when there are multiple variables related to the measurement of a particular element of the model.

An intuitive approach could directly reduce the items related to a certain construct of the model to a single value (using the arithmetic mean, for instance). Another possibility would be to practise aprioristic analyses of the internal consistency of the multi-item scales associated with the constituent constructs of the model, and then identify the most reliable indicators of every multi-item scale with the aim of keeping just one indicator (the best) per construct. The problem 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 us to work with the original format without any pre-processing stage: 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 3.4.2.

3.3 Representation and inclusion of expert 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. In this section, we suggest some alternatives to solve these questions. All of them are based on the expert's ability to express her knowledge in a humanly understandable format by fuzzy logic.

3.3.1 Fuzzy semantics from expert knowledge

First at all, the marketing manager must select the model to be used during the data mining stage. This implies determining both the theoretical constructs and the observed variables/indicators associated with each one, i.e., the measurement model. Therefore, a transformation of the original marketing scales used for measuring those indicators into linguistic terms should be done (fuzzy semantic).

At this point, several marketing scale types can be used for its measurement. With the aim of focusing the problem, we use Stevens (1946, 1959) as a base to summarise them in four, i.e., nominal, ordinal, interval and ratio. Considering those types, a transformation into fuzzy semantic is meaningful for the entire majority with the exception of variables measured by means of a nominal scale, where the nature of categories defining the scale are purely deterministic of the consumer's condition. Furthermore, this transformation should be practised 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, inasmuch as 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 which is the more convenient degree of sensitivity in the fuzzy scales being used in her study. Three or five linguistic terms (fuzzy sets) seem good options.
- The membership function type and shapes defining the behaviour of certain fuzzy variable should be also defined. In this sense, 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 for obtaining a linear behaviour, while Gaussian functions can be used for a non-linear one.
  In this respect, focusing on those marketing scales mainly used for measuring the observed variables related to certain theoretical construct – i.e., Likert, differential semantic or rating scales, which are ordinal scales *sensu stricto*, though assumptions

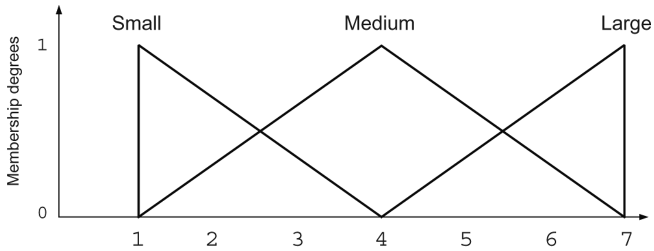
are made to treat them as interval scales, – we pose that it is more appropriate to use linear functions, inasmuch as it facilitates the latter interpretation of relations. Furthermore, we believe a transformation into a triangular function to be more convenient, if special characteristics of these marketing scales are considered – i.e., scales valuations are punctual, then when the membership degree of certain linguistic terms is equal to one, such a term should be associated with a point of the scale.

To sum up, we show an example in Figure 1 based on the transformation of a seven-point rating scale into a three-triangular fuzzy semantic, with the three linguistic terms (small, medium, and large) represented by the corresponding fuzzy sets characterised by the following three membership functions:

$$\mu_{\text{Small}}(x) = \begin{cases} \dfrac{4-x}{3}, & 1 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}, \qquad \mu_{\text{Medium}}(x) = \begin{cases} \dfrac{x-1}{3} & 1 \le x \le 4 \\ \dfrac{7-x}{3}, & 4 < x \le 7 \\ 0 & \text{otherwise} \end{cases},$$

$$\mu_{\text{Large}}(x) = \begin{cases} \dfrac{x-4}{3}, & 4 \le x \le 7 \\ 0, & \text{otherwise} \end{cases}.$$

Figure 1 Transformation of a seven-point rating scale into a three-triangular fuzzy semantic



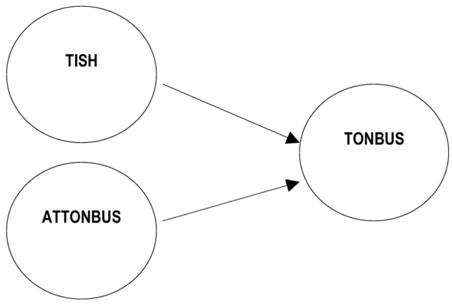
3.3.2 Input/output linguistic variables from expert knowledge

Once the structural model has been fixed by the marketing expert, fuzzy rules are used to relate input (antecedents) to output (consequents) variables. Obviously, hypotheses contained in the structural model can be directly used to define the IF-THEN structures by considering the dependences shown among the variables. Thus, we obtain several fuzzy rule bases for each consequent (endogenous construct) considered and its respective set of antecedents.

For example, given the following structural model (Figure 2) defined by three theoretical constructs or latent variables; i.e.,

- *TISH*: Consumer's overall trust in internet shopping.
- *ATTONBUS*: Consumer's attitude towards certain online business.
- *TONBUS*: Consumer's overall trust in shopping in certain online business.

Figure 2 Example of a structural model



A plausible set of rules obtained by using certain learning method could be the following:

IF *TISH* is {**high** or **medium**} and *ATTONBUS* is **high**
THEN *TONBUS* is **high**
IF *TISH* is **high** and *ATTONBUS* is **medium** THEN
TONBUS is **medium**
IF *TISH* is **small** THEN *TONBUS* is **small**

As can be seen, with this structure, possible non-linear relations among latent variables are analysed and interpreted more easily.

3.4 Machine learning (Data mining process)

Once the linguistic variables that properly represent the information tackled are fixed, a machine learning process must be used to automatically extract the knowledge existing in the considered data. This process is, without any doubt, the most important issue from the knowledge discovery point of view. This section shows a simple and quick learning method based on genetic algorithms (Michalewicz, 1996) to do that. It is designed with the objective of generating *descriptive* fuzzy rules that show some interesting relationships contained in the analysed data.

3.4.1 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 and 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, $i \in \{1,\ldots,n\}$, taking as a value a set of linguistic terms $\tilde{A}_i = \{A_{i1}$ or … or $A_{in_i}\}$, whose members are joined by a disjunctive operator (we use the bounded sum, $\min\{1, a+b\}$, in this paper), whilst the output variables Y_j, $j \in \{1,\ldots,m\}$, remain usual linguistic variables with single labels associated. This structure uses a more compact description that improves the interpretability. Moreover, the structure is a natural support to allow the absence of some input variables in each rule (simply making $\tilde{A}_i$ to be the whole set of linguistic terms available).

3.4.2 Multi-item fuzzification

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

$$\mu_{\tilde{A}_i}(\vec{x}_i) = \max_{h_i \in \{1,\ldots,p_i\}} \min\left\{1, \sum_{l=1}^{n_i} \mu_{A_{il}}(x_{h_i}^i)\right\}$$

while the fuzzy proposition 'Y_j is B_j' is computed as follows:

$$\mu_{B_j}(\vec{y}_j) = \max_{k_j \in \{1,\ldots,m_j\}} \mu_{B_j}(y_{k_j}^j).$$

3.4.3 Discovery process

The subgroup discovery technique (Lavrac et al., 2004) is used in our proposal. It involves running the genetic algorithm once for each possible consequent combination, extracting the most interesting groups (antecedent combination) for each of them. The use of a multi-objective approach allows the algorithm to extract rules with different properties, according to two different criteria as described here.

3.4.4 Multi-objective genetic algorithm

The proposed multi-objective genetic algorithm consists of the following components:

- *Coding scheme*: Each individual of the population represents a single fuzzy rule. Since only the antecedent must be encoded (because the consequent is fixed for each run), a binary representation may be used. Thus, the coding scheme size is 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. For example, assuming we have two input variables and three linguistic terms (small, medium, and large) for each of them, the fuzzy antecedent [IF X_1 is Small and X_2 is {Medium or Large}] is encoded as [100|011].
- *Evolutionary scheme*: A generational approach with the multi-objective NSGA-II replacement strategy (Deb et al., 2002) is considered.
- *Criteria*: We employ two different objectives to extract fuzzy rules with different characteristics:
  - *Support*: This objective function measures the representation degree of the corresponding fuzzy rule among the available data. It is computed by gathering the matching of all the analysed data to the fuzzy rule antecedent as follows:

$$F_1(R) = \text{Support}(R) = \frac{1}{N} \sum_{e=1}^{N} \mu_A(\vec{x}_e),$$

with N being the data set size, $\vec{x}_e = (\vec{x}_1^e, \ldots, \vec{x}_n^e)$ the eth input multi-item data instance and $\mu_A(\vec{x}_e) = \min_{i \in \{1,\ldots,n\}} \mu_{\tilde{A}_i}(\vec{x}_i)$ the matching degree of the rule R for this example (i.e., the fuzzy t-norm *minimum* is used to interpret the connective 'and' of the fuzzy rule).

- *Confidence*: This objective function measures the quality (accuracy) of the relationship between input and output variables described by the analysed fuzzy rule. It is computed as follows:

$$F_2(R) = \text{Confidence}(R) = \frac{1}{N} \sum_{e=1}^{N} \frac{\mu_R(\vec{x}_e, \vec{y}_e)}{\mu_A(\vec{x}_e)},$$

with $\vec{y}_e = (\vec{y}_1^e, \ldots, \vec{y}_m^e)$, $\mu_R(\vec{x}_e, \vec{y}_e) = \mu_A(\vec{x}_e) \cdot \mu_B(\vec{y}_e)$ being the implication result of the fuzzy rule R for the eth data instance $(\vec{x}_e, \vec{y}_e)$, and $\mu_B(\vec{y}_e) = \min_{i \in \{1,\ldots,m\}} \mu_{B_j}(\vec{y}_j^e)$. Therefore, the t-norm *product* is used for the fuzzy implication operator 'THEN.'

- *Selection*: The binary tournament selection based on the crowding distance in the objective function space (NSGA-II) is used.
- *Crossover operator*: A simple two-point crossover operator is considered.
- *Mutation operator*: The mutation randomly selects an input variable of the rule antecedent encoded by the individual. 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).

4 Application of the KDD methodology to a Web-trust consumer's model

The main aim we pursue with this section is to briefly show the potential of the KDD methodology we have theoretically presented in the previous parts of this paper. In this regard, we are aware of the fact that these types of brand new KDD methods do not only need to be well established in theoretical terms. Furthermore, and even more interesting to facilitate the understanding of their benefits, it is desirable and more rigorous to complement such a presentation with an empirical application of the same. This is what we do in this section.

To follow a logical sequence of contents for the reader, we have structured this section in two parts. First, we make some synthetic commentaries about the model and database that we use to show empirical evidence about how this method works. Specifically, we apply the KDD method we present to a Web-trust consumer's model. This, however, is a secondary question, as our primary objective at this point in the paper is not analysing the theoretical basis of the model used, but showing the potential and performance of our methodology. Nevertheless, a brief description of the aim and scope of the model used for the experimentation will aid a better understanding of the utility of this method to extract useful information patterns from an online customers' database.

And, second, we show and discuss a variety of results obtained after applying our descriptive modelling method to the data.

4.1 Previous notes

4.1.1 Application model and data source used for the experimentation

The data employed to empirically illustrate the performance of our KDD method come from a research whose main findings have been published in Martínez-López and Montoro (2004) and Martínez-López et al. (2005). To simplify the experimentation, we have based our application on the less complex model presented in the former reference. The authors have given us permission to use their data and model for our research purposes. In this sense, we would also like to highlight that it is a common practice to employ data and models already existing and validated to apply and analyse the performance of a particular brand new KDD method or, more specifically, a data mining algorithm proposed to support a specific marketing problem (see, as e.g., Beynon et al., 2001; Fish et al., 2004; Hurley et al., 1995; Levy and Yoon, 1995; Rhim and Cooper, 2005). Hence, this option can be seen as methodologically orthodox and rigorous.

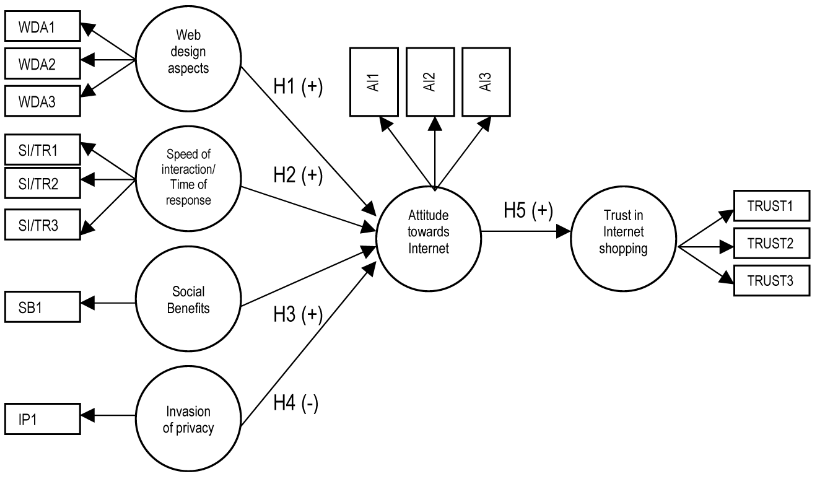
Next, we offer a short explanation of the model use for the experimentation. This is a consumer's trust in internet shopping model based on the standard learning hierarchy, also called CAB – i.e., cognitions → affect → behaviour – paradigm (Holbrook, 1986). The gist of this paradigm is the following (Solomon et al., 2002): the individual (consumer) initially forms beliefs about certain objects by accumulating knowledge with regard to several attributes which define said object (i.e., internet, in this case); then, once these beliefs are developed, feelings – i.e., affective responses – are formed (i.e., attitude towards the internet); lastly, the consumer's behaviour-related responses (i.e., trust in online shopping in the model) are based on those affective responses.

Such organisation of the constituent elements of the model is a plausible way to structure the formation of consumers' attitudes on the electronic markets, as it is adequate to explain high-involvement behaviours like those characterising the online shopping (Wolfinbarger and Gilly, 2001). Basically, the model poses which consumers' beliefs and attitudes towards the internet as a communication medium can be plausible determinants of their trust in internet shopping (see Figure 3). Next, we briefly present each element of the model to facilitate the reader's understanding of them; these are the following:

- *Design aspects*: The consumer's opinion about the availability, design attractiveness and structure of information on the Web in general, and the websites in particular.
- *Interaction speed/time of response*: The consumer's perception about the internet's capacity in general and, more particularly, of websites as a whole, to give a response when required.
- *Social benefits*: The consumer's opinion about the internet's contribution to the well-being of the whole society.

- *Invasion of privacy*: The consumer's opinion regarding the invasion of her intimacy by the various agents with which she interacts in internet applications.
- *Attitude towards the internet*: The consumer's overall evaluation towards the internet communication medium.
- *Trust in internet shopping:* The consumer's perceptions regarding the credibility or reliability of shopping through the internet (the Web).

Figure 3 Theoretical and measurement model used in the experimentation



In the original study, such a theoretical model was estimated by using LISREL for a sample of 529 internet users. The results offered positive results to accept as valid the theoretical structure proposed. In Table 1, we offer a synthesis of the result of the parameterisation process by SEM. Furthermore, such a basic structure was also validated for samples of internet users from different countries (see Martínez-López et al., 2005).

The measurement issues for each of the six elements/constructs of this model were adequately treated in the original study. In this regard, most part of the constructs were measured by using scales already validated in previous studies; specifically, all of them were based on seven-points Likert-type and differential semantic scales; in Figure 3, we also graphically show the measurement model we work with. In any case, the measurement scales reliability analyses realised by the authors offered satisfactory results.

In sum, these are some key questions that allow us to rely on these data and model to offer a good empirical illustration of our KDD method based on them. Likewise, considering the scope of the model, the reader should better see how the kind of information this method provides may improve the view and support of the marketing manager's decision-making on the Web-based markets.

Table 1 Estimation by SEM of the model used for the experimentation

	Relations between constructs in the model	*Hypothesised sign*	*Standard parameters*
Hypothesis 1	Web design aspects → Attitude toward internet	positive	0.38***
Hypothesis 2	Interaction speed/time of response → Attitude toward internet	positive	0.11*
Hypothesis 3	Social benefits → Attitude toward internet	positive	0.36***
Hypothesis 4	Invasion of privacy → Attitude toward internet	negative	–0.15*
Hypothesis 5	Attitude toward internet → Trust in internet shopping	positive	0.44**

*Coefficient significant at $p < 0.05$.

**Coefficient significant at $p < 0.01$.

***Coefficient significant at $p < 0.001$.

4.1.2 Conversion of the web-trust model into fuzzy systems

With respect to the transformation of the model into fuzzy systems, it contains two endogenous elements/variables; i.e., attitude towards the internet and trust in internet shopping. As is widely known, the term 'endogenous' refers to an element of any causal model which comes defined by others. Therefore, two Fuzzy Rule Sets (FRS) will be obtained to explain the two endogenous concepts. The former set (FRS1) will contain rules where the consequent is "attitude towards the internet" and the four beliefs are considered as antecedent, while the latter one (FRS2) 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 to focus on the more relevant fuzzy rules to extract information about the consumer behaviour being modelled.

4.2 Results

As previously mentioned, in this section we offer an illustrative view of the performance of our KDD method. In this regard, though we select and analyse some rules, from the whole sets of rules generated by the data mining stage, in terms of their confidence and support, the reader should realise that other rules could be taken into account. In other words, the KDD process may also present certain degrees of subjectivity, in this case introduced by the marketing expert when post-processing the output of the machine learning stage. Specifically, the selection of rules is usually based on both objective questions like the quality metrics of support and confidence and the interest of the marketing expert in achieving certain information (mostly related with the decision to be taken).

We have structured this section in two main parts, which is defined by every FRS we work with; i.e., the endogenous variables of the model. In particular, each of these sections shares the following structure:

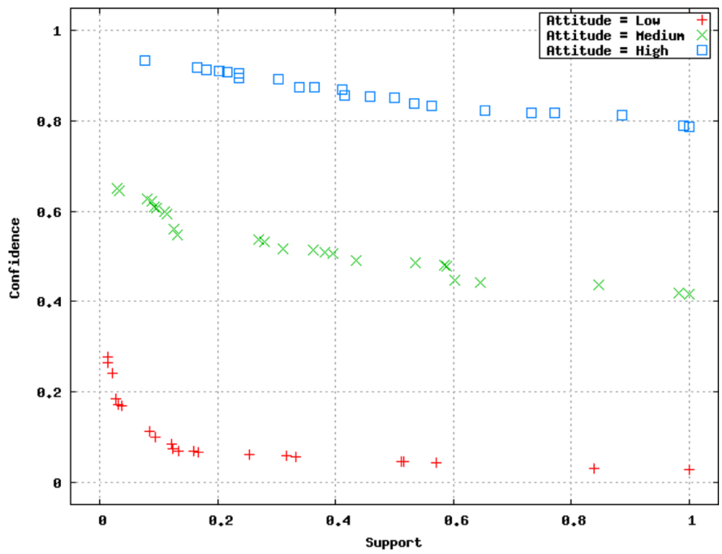
- *Overall analysis of the Pareto Front*. This resource offers visual information regarding the maximum frontier to achieve, so the best/non-dominated rules, in terms of support and confidence. Likewise, the algorithm has been designed not only to search and generate the best rules in absolute terms, 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.
- Selection of descriptive rules and interpretation.
- Useful knowledge to support the manager's decisions addressed to the online consumer.

4.2.1 Attitude towards the internet

4.2.1.1 Pareto front analysis for attitude towards the internet

The information provided by this Pareto (see Figure 4) is quite clear regarding the behaviour of the variable of interest, in this case; i.e., attitude towards the internet.

Figure 4 Pareto front: design and speed and social and privacy vs. attitude (see online version for colours)



On the one hand, the absolute Pareto front is formed, exclusively, by rules whose consequent is high. Besides, 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 till the rule with higher, with a maximum and minimum around 0.95 and 0.80, respectively. This means that there is a significant set of rules with notable support – thus information patterns with wide representation on the consumers' database – whose consequent is high.

On the other hand, we have obtained two sub-Pareto fronts whose rules generally present non-reliable levels of confidence; i.e., those associated with low and moderate levels of the consumers' attitude towards the internet. This fact is especially evident for the rules whose consequent is low.

In sum, we can extract a general idea, based on this visual information: the consumers' attitude towards the internet is mostly high, regardless the beliefs they may have towards this medium. However, they can present moderate levels of attitude in certain non-generalised cases (rules with low support), though it loses significance as the support of the rules increases. In any case, the degree of confidence in this consequent for attitude is poor, which reinforces the starting idea of this paragraph.

4.2.1.2 Selection of descriptive rules and interpretation

As the reader should have rapidly seen after having a look at the Pareto, the fuzzy rules set associated with attitude towards the internet is compounded by a considerable number of rules. Specifically, if we consider the whole set of rules for the three sub-Pareto fronts, there are over 70; i.e., the total of plots in Figure 4. However, the task of selection is easier since we know that some of them do not offer reliable information. Usually, the marketing expert should focus on the constituent rules of the absolute Pareto front; i.e., the best rules. In this case, there are 22 rules.

To show this part of our methodology, we next analyse some interesting rules from a selection made considering their levels of reliability and support (see Table 2); these are rules belonging to the absolute front. In any case, we highlight again our illustrative purposes. This means that a marketing expert could also select and analyse others depending on his information needs.

Table 2 Fuzzy rules extracted from the absolute front for beliefs and attitude towards internet

	Fuzzy rules	*Support*	*No. cases*	*Confidence*
[1]	IF Design is **high** and Speed is **high** and Social is **high** and Privacy is **low** THEN Attitude is **HIGH**	0.0767	78	0.9351
[2]	IF Design is **high** and Speed is {**medium** or **high**} and Social is **high** and Privacy is **low** THEN Attitude is **HIGH**	0.1648	159	0.9195
[3]	IF Design is **high** and Social is **high** and Privacy is **low** THEN Attitude is **HIGH**	0.1805	167	0.9146
[4]	IF Design is **high** and Social is **high** THEN Attitude is **HIGH**	0.3027	264	0.8920
[5]	IF Social is **high** THEN Attitude is **HIGH**	0.4115	306	0.8687
[6]	IF Design is **high** THEN Attitude is **HIGH**	0.5344	397	0.8395
[7]	Regardless of the antecedents (beliefs) Attitude is **HIGH**	1	469	0.7867

The first rule we analyse, the rule 1, is the most reliable information pattern in the whole set of rules generated. It represents a particular scenario where the four beliefs take the kind of values that hypothetically generate the best opinions towards internet. Certainly, though its support is very low, what it means is that this 'optimum' scenario is not usual among online consumers, it is very accurate. This allows us to affirm that it is very reasonable to expect that a consumer must show good attitudes towards the internet, when she presents such kinds of beliefs. However, the rules we have obtained do not allow asseverating the contrary. Specifically, if we go through the whole set of rules and searched the opposite rule in terms of beliefs and attitude, we would find a rule with minimum support (0.01) and a confidence lower than 0.3. Therefore, such a rule clearly says that it is not plausible that consumers have low levels of attitude towards the internet, even when they present a configuration of beliefs which supposedly produce it, as has been hypothesised. We think that this is a very interesting example of the kind of information we can achieve with this method.

Next, we focus on rules 2 and 3. First, the reader can see how rule 2 is similar to rule 1, with just one variation in the value taken by the consumer's perception about the speed of interaction. In this case, if the consumer presented worse opinions (i.e., moderate) about the speed of interaction, ceteris paribus, her level of attitude towards the internet would remain high. This information leads us to think that this belief is not very influential in determining high levels of attitude. In this regard, if we have a look at rule 3, we can confirm this idea. In essence, this rule says that, whatever the consumer's perception about the speed of interaction is, when the rest of the beliefs maintain their values, her attitude stays high.

Other interesting rules are 4–6. We analyse them all together, as they allow the extraction of relevant information about the relations between beliefs and attitude. These rules still have acceptable levels of reliability. This means that, though the confidence of the information patterns they present is not as good as the previous rules, probably as a consequence of the increase in the support of the rules, it is still reliable. Once we know we are working with a trustworthy information pattern, we are going to use these rules to achieve the following conclusion: the beliefs design and social benefits are truly influential in determining high levels of attitude, the other two being marginal or non-significant. This question easily arouses the suspicion of a 'trained eye' (i.e., the individual who is used to working with these kinds of rules) when having a look at the ruled generated. Specifically, rule 4 basically says that, regardless of the consumer's beliefs about speed and privacy, if her opinions about the design aspects and the social benefits provided by the internet are high, then attitude should be good; this rule has a good confidence, around 0.9, as well as a respectable support, which suggests that it is not an unusual case. In other words, it seems that design and social are the beliefs which are truly influential in determining high levels of attitude. The relevance of these beliefs is also suggested by rules 5 and 6, where we can see how each of these two factors could generate, by themselves, high levels of attitude. However, the confidence of these rules is not as good as the reliability of rule 4. Thus, this question recommends that both factors be jointly considered when analysing their influence on attitude towards the internet.

Finally, from the total of rules integrating the absolute Pareto front, we have also selected a particular rule we would like to comment on. Specifically, rule 7 says that no matter what the consumer's beliefs about the internet are, her attitude is always high. In other words, this information pattern suggests that the set of beliefs considered in the model does not produce any effect on attitude. However, its confidence is not good. But, the reader should think how useful this rule would be if it had good reliability. This would imply getting a clear and simple conclusion about the non-influence of the set of beliefs. Notwithstanding, it is not the case.

Apart from the information we have commented on that an expert may extract from a selection of rules belonging to the absolute Pareto front, there may be other interesting rules which could deserve consideration. We refer to those rules out of the absolute front that may also present good levels of confidence or, on the contrary, rules with poor reliability which may be used to reject the information patterns they show. Next, we illustrate this with our empirical application.

As we can see in the Pareto, there are no rules with good reliability out of the absolute front; they have confidence levels of around 0.6 or lower. Thus, we cannot play with this possibility, as such information patterns are not trustworthy. However, we have selected two rules (see Table 3) which are of help to precisely the contrary, i.e. to confirm that a certain scenario is not plausible. For instance, rule 8 represents the opposite situation if compared with rule 1, what we called the optimum scenario. In this case, rule 8 shows the configuration of beliefs that should supposedly produce, as hypothesised, the lowest attitude. But, if the confidence associated with rule 8 is checked, it is clear that this information pattern is not reliable. So, we can conclude, with no risk, that even when the online consumers present such a configuration in their beliefs about internet, their attitude towards this medium will not be low. Furthermore, if we also use rule 9, we could say that it would be more plausible that a consumer had moderate than low levels of attitude.

Table 3 Examples of fuzzy rules not belonging to the absolute front for beliefs and attitude towards internet

Fuzzy rules	*Support*	*No. cases*	*Confidence*
[8] IF Design is **low** and Speed is **low** and Social is **low** and Privacy is **high** THEN Attitude is ***LOW***	0.0127	13	0.2777
[9] IF Design is **low** and Speed is **medium** and Social is **low** and Privacy is **high** THEN Attitude is ***MEDIUM***	0.0092	11	0.6666

4.2.1.3 Useful knowledge for the marketing manager

What can be concluded from all the above? On the one hand, consumers do not really have a poor attitude towards the internet. Thus, even when opinions about the different beliefs considered are not good, it is more likely that consumers present a moderate attitude. In any case, on the basis of our Web users' database, it should also be highlighted that though a part of the consumers have a moderate attitude towards the internet, the majority of them show good levels of attitude.

The set of relations between beliefs and attitude already supported by means of SEM seems to be verified when analysing the first fuzzy rule set. However, due to low attitudes being very unusual, beliefs are really influential in determining attitude variations from moderate to high levels. These are the kind of questions that we can discover with our method, which are difficult to see by means of SEM. In this sense, *design of websites* and the *social benefits* provided by the internet are the most influential beliefs to produce good levels of attitude in consumers, while the other two beliefs are clearly non-influential.

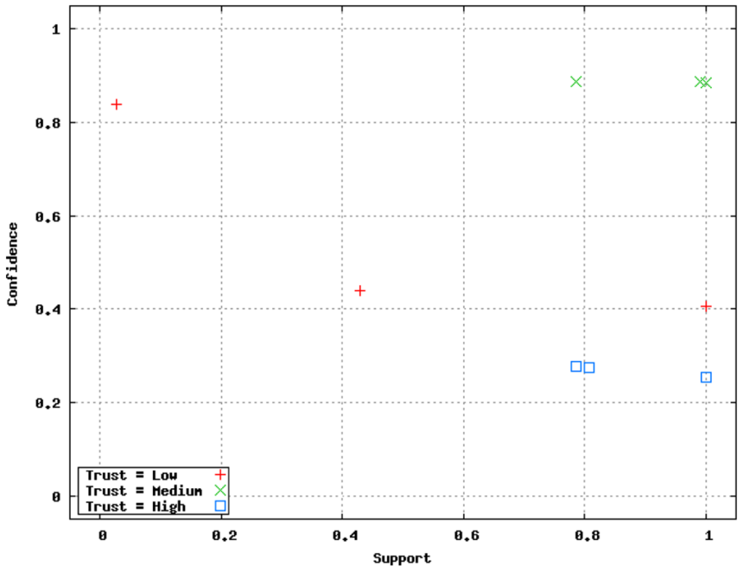
4.2.2 Trust in internet shopping

4.2.2.1 Pareto front

The simplicity of this system, which represents the relation between just one antecedent (attitude towards the internet) and a consequent (trust in internet shopping), is obvious when taking a look at its Pareto front (see Figure 5). The total of rules integrating the Pareto fronts is not large. For instance, if we compare this system with the previous, much more complex, the difference is enormous: 14 rules for the

FRS2 vs. more than 70 for the FRS1. This makes the analysis of the rules easier, especially if we take into account that the number of rules finally kept to be analysed is smaller.

Figure 5 Pareto front: attitude vs. trust in Web shopping (see online version for colours)



The absolute Pareto front in this case is formed by four rules; one whose consequent is low and three whose consequent is medium. Likewise, there are no significant rules associated with high levels of trust, either with low or high support of the consumers' database. In sum, based on this visual information, we can conclude that the range of variation of the consumers' trust in internet shopping moves from low to medium, though the consumers show, in general, moderate levels of trust.

4.2.2.2 Selection of descriptive rules and interpretation

In Table 4, we show a significant selection of rules from the absolute Pareto front belonging to the FRS2. First, we focus on rule 10. This is a very useful rule as it covers nearly the whole database, and its confidence is close to 0.9; it is one of the most reliable rules from the whole FRS2. Specifically, this rule says that it is very likely that when the consumers' attitude towards the internet is moderate or good, her trust in shopping through the Web is moderate. This is interesting information as it suggests that, in the case of influence, attitude towards the internet does not produce a constant effect on trusting online shopping. Likewise, rule 11 is also valuable. As we have already commented, it is unusual that consumers show low levels of attitude towards the internet (support around 0.02 and 20 cases). However, it is very probable that if the consumers' attitude is poor, then their trust in internet shopping will be low (confidence around 0.83).

Table 4 Fuzzy rules extracted from the absolute front for attitude towards internet vs. trust in internet shopping

Fuzzy rules	*Support*	*No. cases*	*Confidence*
[10] IF Attitude is {**medium** or **high**} THEN Trust is ***MEDIUM***	0.990050	468	0.8870
[11] IF Attitude is **low** THEN Trust is ***LOW***	0.026297	20	0.8378

Besides, we could complement the previous analysis, going to the rest of the rules from the absolute Pareto front. Specifically, we highlight one rule, the one which relates high levels of attitude with high levels of trust. As the reader can imagine, it would not be logical that this rule presents good confidence. This fact would be in contradiction to rule 10. Indeed, that new information pattern we comment on has poor reliability, with confidence of around 0.3. In essence, this reinforces some ideas we could extract from the analysis of the FRS2:

- it is atypical to find good levels of consumers' trust in internet shopping
- attitude is not significant in causing high levels of trust.

4.2.2.3 Useful knowledge for the marketing manager

The results seem to verify the relation between attitude and trust tested by SEM. However, we would like to give several important comments, which clearly add qualitative information for the manager. Firstly, most consumers have a moderate opinion about trusting in internet shopping. Secondly, attitude shows a really weak influence on trusting in internet shopping when that varies from moderate to high levels. So, it is not appropriate to conclude that the intensity of the relation keeps both linear and invariable, as the linear coefficient obtained by SEM supposes, inasmuch as if attitude is moderate or high, it is very plausible that trust is moderate. Finally, attitude exerts a big influence on trusting when that varies from low to moderate levels.

5 Concluding remarks

We have attempted to make one more step in the rapprochement of the KDD methodologies and the artificial intelligence discipline to the field of the decision-support systems in marketing. Specifically, our research has paid especial attention to discussing the diversity of questions we tackle with a holistic perspective, trying to jointly consider the academic and the professional arena.

Our research has interesting insights for marketing academics and practitioners with regard to the main and the specific topic we deal with. We first introduce a general framework of discussion. Firms are increasingly concerned about the transcendence of having timely and good quality information to guide their decisional processes on consumer markets. As information technologies have evolved during the last decades, companies have found fewer difficulties in accessing consumer data. Specifically, the Web-based markets (i.e., the electronic markets), characterised for being information-intense environments, are spaces where accessing and collecting data is not especially difficult, if firms own the right tools. Therefore, in the present day, having data does not imply by itself a key to determine the competitiveness of firms. On the contrary, one of the factors which is definitely essential to achieve entrepreneurial success is the use that firms make of such data. In other words, businesses must care about owning suitable capabilities to transform such data into high value information. This is, as discussed in the introductory section, an important challenge to be faced by firms.

Consequently, it is not strange that one change in the use which firms make of MkMSS, as seen in the last decade, is a shift from data-driven MkMSS to knowledge-based MkMSS. In this sense, the analytical capabilities based on artificial intelligence methodologies have demonstrated that they are very useful for guiding the marketing managers' decisional processes on consumer markets. This is why the KDD methods have been more intensively applied to solve information problems on markets. However, as we have noted, the reader should realise that these methods, as well as the technological advances adapted from the artificial intelligence discipline for the analysis of data, are not a panacea. On the contrary, first we should be aware of the fact that firms have to deal with different marketing problems, whose resolution may take diverse ways. In this regard, it is logical to think that marketing managers should make an oriented use of the analytical tools available, either belonging to the statistical techniques or to the artificial intelligence methods. Specifically, every data mining paradigm (as e.g., rules induction, artificial neural networks, decision tress, evolutionary algorithms, etc.). or KDD method, in general, has its own distinguishing characteristics which make it suitable to be applied for solving certain information problem. And, second, we suggest the application of such KDD methods with a cooperative and complementary spirit to the information the managers can also obtain by means of certain statistical techniques.

In particular, the main contribution of this research is the proposal and empirical experimentation of a complete KDD methodology based on genetic-fuzzy systems. Such methodology has been designed with the aim of being applied to consumer behaviour modelling. Specifically, it is able to work with the kind of variables and measurement models which are usually designed by the marketing modellers. Till now, SEM was the only method able to estimate such complex causal models of consumer behaviour. However, as we have discussed in depth, such a technique, though useful for academic purposes, does not perform well when applied as a decision-support tool in real market situations. The information provided is deficient, so it needs to be complemented somehow. Otherwise, the marketing manager may perceive too much uncertainty when taking the decisions. With this purpose, our method successfully hybridises fuzzy logic with genetic algorithms to extract useful and descriptive information patterns from customers' databases, prior to the application of a data mining process supervised by a model of reference. Then, we can be assured of working with optimum solutions, expressed in an easy, semantically understandable way of reasoning of the human being.

The results we have obtained prove the viability of the method we propose, as well as its potential when applied to better understanding of the behaviour of consumers. Nevertheless, there is still a long way to go, with very interesting research streams to be developed, in the fascinating and necessary task of improving the analytical capabilities of marketing decision-support systems, by importing and adapting KDD and artificial intelligence-based methods.

Acknowledgement

Research supported in part by the Spanish Ministry of Education and Science, under project No. TIN2005-08386-C05-01.

References

Aziz, N.A. and Yasin, N.M. (2004) 'The influence of market orientation on marketing competency and the effect of Internet-marketing integration', *Asia Pacific Journal of Marketing and Logistics*, Vol. 16, No. 2, pp.3–19.

Baumgartner, H. and Homburg, C. (1996) 'Applications of structural equations modelling in marketing and consumer research: a review', *International Journal of Research in Marketing*, Vol. 13, pp.139–161.

Berthon, P. and John, J. (2006) 'The marketing-IT paradox: interactions form the customer's perspective', in Sheth, J.N. and Sisodia, R.S. (Eds.): *Does Marketing Need Reform? Fresh Perspectives on the Future*, M.E. Sharpe, NY, pp.191–197.

Beynon, M., Curry, B. and Morgan, P. (2001) 'Knowledge discovery in marketing: an approach through rough set theory', *European Journal of Marketing*, Vol. 35, Nos. 7–8, pp.915–935.

Bollen, K.A. and Long, J.S. (Eds.) (1993) 'Introduction', *Testing Structural Equation Models*, Sage Publications, Inc., pp.1–9.

Carlsson, C. and Turban, E. (2002) 'Introduction. DSS: directions for the next decade', *Decision Support Systems*, Vol. 33, pp.105–110.

Casillas, J., Cordón, O., Herrera, F. and Magdalena, L. (Eds.) (2003a) *Interpretability Issues in Fuzzy Modeling*, Springer, Heidelberg, Germany.

Casillas, J., Cordón, O., Herrera, F. and Magdalena, L. (Eds.) (2003b) *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, Heidelberg, Germany.

Cravens, D.W., Lamb, C.W. and Crittenden, V.L. (2002) *Strategic Marketing Management*, McGraw-Hill, Boston.

Deb, K., Pratap, A., Agarwal, S. and Meyarevian, T. (2002) 'A fast and elitist multiobjective genetic algorithm: NSGA-II', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp.182–197.

Ein-Dor, P. and Jones, C.R. (1985) *Information Systems Management: Analytical tools and Techniques*, Elsevier, New York.

Evans, P.B. and Wurster, T.S. (1997) 'Strategy and the new economics of information', *Harvard Business Review*, Vol. 75, No. 5, September–October, pp.71–82.

Fayyad, U.M., Piatesky-Shapiro, G., Smyth, S. and Uthurusamy, R. (1996) *Advances in Knowledge Discovery and Data Mining*, M.I.T. Press, Cambridge, Massachusetts.

Fish, K.E., Johnson, J.D., Dorsey, R.E. and Blodgett, J.G. (2004) 'Using an artificial neural network trained with a genetic algorithm to model brand share', *Journal of Business Research*, Vol. 57, No. 1, pp.79–85.

Freitas, A.A. (2002) *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer, Heidelberg, Germany.

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*, Vol. 17, pp.209–214.

González, A. and Pérez, R. (1998) 'Completeness and consistency conditions for learning fuzzy rules', *Fuzzy Sets and Systems*, Vol. 96, No. 1, pp.37–51.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998) *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ.

Hannula, M. and Pirttimaki, V. (2003) 'Business intelligence empirical study on the top 50 Finnish companies', *Journal of American Academy of Business*, Vol. 2, No. 2, pp.593–599.

Heiens, R.A. (2000) 'Market orientation: toward an integrated framework', *Academy of Marketing Science Review*, No. 1, Available at: http://www.amsreview.org/articles/heiens01-2000.pdf

Hoffman, D. and Novak, T. (1996) 'Marketing in hypermedia computer-mediated environments: conceptual foundations', *Journal of Marketing,* Vol. 60, No. 3, July, pp.50–68.

Holbrook, M.B. (1986) 'Emotion in the consumption experience: toward a new model of consumer behavior', in Peterson, R.A., Hoyer, W.D. and Wilson, W.R. (Eds.): *The Role of Affect in Consumer Behavior*, Lexington Books, Lexington, MA, pp.17–52.

Hunt, S.D. and Morgan, R.M. (1995) 'The comparative advantage theory of competition', *Journal of Marketing*, Vol. 59, No. 2, April, pp.1–15.

Hurley, S., Moutinho, L. and Stephens, N.M. (1995) 'Solving marketing optimization problems using genetic algorithms', *European Journal of Marketing*, Vol. 29, No. 4, pp.39–56.

Jöreskog, K.G. and Sörbom, D. (1993) *LISREL 8: Structural Equation Modelling with the SIMPLIS™ Command Language*, Scientific Software International, Chicago, IL, USA.

Kohli, A.K. and Jaworski, B.J. (1990) 'Market orientation: the construct, research propositions, and managerial implications', *Journal of Marketing*, Vol. 54, No. 2, April, pp.1–18.

Kumar, V. and Ramani, G. (2006) 'Interaction orientation: the new marketing competency', in Sheth, J.N. and Sisodia, R.S. (Eds.): *Does Marketing Need Reform? Fresh Perspectives on the Future*, M.E. Sharpe, NY, pp.109–118.

Lan, P. (2006) 'A framework for innovation decision making in the internet age', *International Journal of Management and Decision Making*, Vol. 7, No. 1, pp.105–118.

Laplante, B., Sabourin, S., Cournoyer, L.G. and Wright, J. (1998) 'Estimating nonlinear effects using a structured means intercept approach', in Rigdon, E.E., Schumacker, R.E. and Wothke, W. (Eds.): *A Comparative Review of Interaction and Nonlinear Modelling*, Lawrence Erlbaum Associates, pp.183–202, Mahwah, NJ, USA.

Laudon, K. C. and Laudon, J. P. (1998) *Management Information Systems: New Approaches to Organization and Technology*, Prentice-Hall, Upper Saddle River, NY.

Laurent, G. (2000) 'Improving the external validity of marketing models: a plea for more qualitative input', *International Journal of Research in Marketing*, Vol. 17, pp.177–182.

Lavrac, N., Cestnik, B., Gamberger, D. and Flach, P. (2004) 'Decision support through subgroup discovery: three case studies and the lessons learned', *Machine Learning*, Vol. 57, Nos. 1–2, pp.115–143.

Levy, J.B. and Yoon, E. (1995) 'Modeling global market entry decision by fuzzy logic with an application to country risk assessment', *European Journal of Operational Research*, Vol. 82, pp.53–78.

Li, S., Kinman, R., Duan, Y. and Edwards, J.S. (2000) 'Computer-based support for marketing strategy development', *European Journal of Marketing*, Vol. 34, Nos. 5–6, pp.551–575.

Lindskog, P. (1997) 'Fuzzy identification from a grey box modeling point of view', in Hellendoorn, H. and Driankov, D. (Eds.): *Fuzzy Model Identification*, Springer-Verlag, Heidelberg, Germany, pp 3–50.

Martínez-López, F.J. and Montoro, F.J. (2004) 'Modelling consumer's trust in internet shopping based on the standard learning hierarchy: a structural approach', *Journal of Internet Business*, No. 2, April, Available at: http://bizfac.com/jib/iss02_lopez.pdf

Martínez-López, F.J., Luna, P. and Martínez, F.J. (2005) 'Online shopping, the Standard learning hierarchy, and consumers' internet expertise: an American-Spanish comparison', *Internet Research*, Vol. 15, No 3, pp.312–334.

Martínez-López, F.J., Montoro, F.J. and Maraver, G. (2006) 'Causal modelling in marketing research: a descriptive review of the testing techniques used in the last decade', *Proceedings of the 35th European Marketing Academy Conference*, Athens, Greece.

Metaxiotis, K., Psarras, J.E. and Samoulidis, J.E. (2004) 'New applications of fuzzy logic in decision support systems', *International Journal of Management and Decision Making*, Vol. 5, No. 1, pp.47–58.

Michalewicz, Z. (1996) *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Heidelberg, Germany.

Mueller, A.R. (1996) *Basic Principles of Structural Equation Modelling: An Introduction to Lisrel and Eqs*., Springer.

Narver, J.C. and Slater, S.F. (1990) 'The effect of a market orientation on business profitability', *Journal of Marketing*, Vol. 54, October, pp.20–35.

Rhim, H. and 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*, Vol. 22, pp.159–182.

Roberts, J.H. (2000) 'The intersection modelling potential and practice', *International Journal of Research in Marketing*, Vol. 17, pp.127–134.

Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R. and Carlsson, C. (2002) 'Past, present and future of decision support technology', *Decision Support Systems*, Vol. 33, pp.111–126.

Silvi, R. and Cuganesan, S. (2005) 'Investigating the management of knowledge for competitive advantage: a strategic cost management perspective', *Journal of Intellectual Capital*, Vol. 7, No. 3, pp.309–323.

Smith, A.D. (2005) 'Exploring online dating and customer relationship management', *Online Information Review*, Vol. 29, No. 1, pp.18–33.

Solomon, M., Bermossy, G. and Askegaard, S. (2002) *Consumer Behaviour: A European Perspective*, 2nd ed., Prentice Hall.

Steenkamp, J. and Baumgartner, H. (2000) 'On the use of structural equation models for marketing modelling', *International Journal of Research in Marketing*, Vol. 17, pp.195–202.

Stevens, S.S. (1946) 'On the theory of scales of measurement', *Science*, pp.677–680.

Stevens, S.S. (1959) 'Measurement, psychophysics and utility', Chap. 2, in Churchman, C.W. and Ratoosh, P. (Eds.): *Measurement: Definitions and Theories*, John Wiley, New York.

Talvinen, J.M. (1995) 'Information systems in marketing: identifying opportunities for new applications', *European Journal of Marketing*, Vol. 29, No. 1, pp.8–26.

Tate, R.L. (1998) 'Effect decomposition in interaction and nonlinear models', in Rigdon, E.E., Schumacker, R.E. and Wothke, W. (Eds.): *A Comparative Review of Interaction and Nonlinear Modelling*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp.167–192.

van Bruggen, G.H. and Wierenga, B. (2000) 'Broadening the perspective on marketing decision models', *International Journal of Research in Marketing*, Vol. 17, pp.159–168.

Vargo, S.L. and Lusch, R.F. (2004) 'Evolving to a new dominant logic for marketing', *Journal of Marketing*, Vol. 68, January, pp.1–17.

Wedel, M., Kamakura, W. and Böckenholt, U. (2000) 'Marketing data, models and decisions', *International Journal of Research in Marketing*, Vol. 17 (Special issue on marketing modelling on the threshold of the 21st century), pp.203–208.

Wierenga, B. and van Bruggen, G.T. (2000) *Marketing Management Support Systems: Principles, Tools and Implementation*, Kluwer Academic Publishers, Norwell, Massachusetts, USA.

Witten, I.H. and Frank, E. (1999) *Tools for Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kauffman Publishers, San Francisco, CA, USA.

Witten, I.H. and Frank, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, San Francisco, CA, USA.

Wolfinbarger, M. and Gilly, M. (2001) 'Shopping online for freedom, control, and fun', *California Management Review*, Vol. 43, No. 2, pp.35–55.

Wood, E. (2001) 'Marketing information systems in tourism and hospitality small and medium- sized enterprises: a study of internet use for market intelligence', *International Journal of Tourism Research*, Vol. 3, No. 3, pp.283–299.

Zadeh, L.A. (1965) 'Fuzzy sets', *Information and Control*, Vol. 8, pp.338–353.

Zadeh, L.A. (1975). 'The concept of a linguistic variable and its application to approximate reasoning. Parts I, II and III', *Information Science*, Vol. 8, pp.199–249, 301–357, Vol. 9, pp.43–80.