

# Obținerea de informații prin aplicarea sistemelor de mulțimi aleatorii, aplicată la modelarea comportamentului consumatorului

Jorge CASILLAS  
Francisco J. MARTÍNEZ-LÓPEZ  
Universitatea Granada, Spania

## Rezumat

*Disciplina care studiază comportamentul consumatorului folosește, prin tradiție, modele care ajută la înțelegerea consumatorilor. Astfel, urmând metoda științifică, reprezentanții mediului academic de marketing propun modele teoretice, ulterior verificate prin intermediul unor metode statistice. Atunci când aceste modele sunt complexe – adică includ construcții dependente și independente legate între ele prin relații multiple – metoda folosită de obicei pentru estimare este Modelarea Ecuatiei Structurale (SEM). În acest sens, hegemonia SEM în estimarea acestui tip de modele de consum este evidentă în ultimele decenii. Totuși, suntem de părere că această metodă are anumite lacune care îi limitează utilitatea în afara cadrului academic; mai precis, testarea modelelor este utilă, deși rezultatele oferite de SEM nu sunt suficient de bune pentru a constitui baza necesară managerilor de marketing pentru luarea deciziilor de piață.*

*Așadar, principala motivație a acestei lucrări se bazează pe convingerea noastră că este necesară o evoluție a metodelor analitice de modelarea în marketing, considerându-se că aplicarea altor instrumente de analiză este mai potrivită în luarea deciziei de către managerii de marketing.*

*Lucrarea de față prezintă pe scurt o metodologie cu totul nouă ce poate fi aplicată în modelarea (cauzală) în marketing. Mai precis, aplicăm experimental această metodă asupra unui model de comportament al consumatorului. Ținând cont de caracteristicile problemei (cu date incerte și informații disponibile de la un expert în marketing) și de optimizarea cu obiective multiple pe care o propunem, sistemele aleatorii genetice sunt un instrument util. Pe scurt, prin aplicarea acestei metodologii se obțin tipare de informații utile (reguli aleatorii) care ajută la o înțelegere mai bună a relațiilor dintre elementele sistemului de marketing analizat (modelul causal); în cazul nostru, un model centrat pe consumator.*

**Cuvinte cheie:** modelare de marketing, sisteme de suport a deciziei, metodologia obținerii de informații, sisteme de mulțimi aleatorii, modele ale comportamentului consumatorului

# Knowledge Discovery by Genetic Fuzzy Systems Applied to Consumer Behavior Modelling

Jorge CASILLAS  
Francisco J. MARTÍNEZ-LÓPEZ  
University of Granada, Spain

## Abstract

*Consumer behaviour discipline has made traditionally use of models to understand consumers. Thus, following the scientific method, marketing academics usually pose theoretical models which are subsequently tested by means of several statistical methods. When such models are complex –i.e. several dependent and independent constructs with multiple relations among them– the method usually used for estimating it is Structural Equation Modelling (SEM). In this sense, the hegemony of SEM for estimating this kind of consumer models has been quite obvious during the last decades. However, we think that this method presents some lacks which constraints its usefulness beyond an academic framework; i.e. it is useful to test models, though results provided by SEM are not good enough for being the kind of support that marketing managers need for guiding their market decisions.*

*Thus, the main motivation of this paper is caused by a strong belief in the necessity that marketing modelling analytical methods have to evolve, considering the application of other tools of analysis more appropriate to aid the marketing managers' decisional processes.*

*This paper briefly presents a brand new methodology to be applied in marketing (causal) modeling. Specifically, we apply it to a consumer behavior model used for the experimentation. The characteristics of the problem (with uncertain data and available knowledge from a marketing expert) and the multiobjective optimization we propose make genetic fuzzy systems a good tool for tackling it. In sum, by applying this methodology we obtain useful information patterns (fuzzy rules) which help to better understand the relations among the elements of the marketing system (causal model) being analyzed; in our case, a consumer model.*

**Keywords:** marketing modeling, decisions support systems, knowledge discovery methodology, genetic fuzzy systems, consumer's behavior patterns.

## 1. Generalități

Reprezentanții mediului academic și practicienii au scos în evidență nevoia de a cunoaște și de a explica, într-un mod tot mai eficient, tiparele comportamentale ale consumatorilor. Firmele orientate spre piețele finale sunt implicate în sisteme înalt competitive în care este necesar ca procesul de luare a deciziei să fie cât mai corect.

Astfel, fiind modele de marketing, modelele de comportament al consumatorului sunt considerate drept un caz specific de sistem de suport al managementului de marketing (MkMSS) și, în timp, și-a demonstrat valoarea de sursă relevantă pentru dezvoltarea științei marketingului (van Bruggen & Vierenga, 2000).

Totuși, modelele actuale de comportament al consumatorului nu par să acopere toate cerințele pe care ar trebui să le satisfacă un model al cărui scop este acela de a ajuta în luarea deciziilor de marketing. În acest sens, pe baza afirmațiilor lui Gatignon (2000), ținând cont de aspectele teoretice și tehnice care încearcă să explice luarea deciziei de către consumatori, va trebui ca modelele viitoare să funcționeze corespunzător cerințelor utilizatorilor față de aceste modele (elementul cerere). Adică, modelele trebuie să fie mai complete, flexibile și structurate în funcție de specificul strategic al mediului competitiv în care operează utilizatorii lor. Astfel, cum principala problemă cu care se confruntă de fapt firmele orientate spre piața consumatorilor nu este disponibilitatea informațiilor (datelor), ci deținerea nivelului de cunoștințe necesar pentru a lua deciziile potrivite, folosirea modelelor comportamentale de avangardă, capabile să pună în valoare aceste informații, poate reprezenta o sursă esențială de avantaj competitiv.

Fără îndoială, este timpul să regândim rolul cercetării de marketing și structurarea de modele, pentru a dezvolta metode analitice mai adecvate, potrivite pentru studiul mediilor de afaceri actuale și a scenariilor decizionale (Wind, 2006). Se așteaptă ca MkMSS să-și îmbunătățească performanța, profitând de sinergiile determinate de integrarea tehnicilor de estimare a modelelor pe baza econometriei clasice, cu ajutorul unor sisteme sofisticate, bazate pe inteligența artificială.

Mai precis, luând în considerare cei trei piloni pe care se bazează modelarea în marketing (Roberts, 2000) și în special modelarea comportamentului consumatorilor, lucrarea de față se ocupă de unul dintre aceștia: tehnicile de estimare a modelelor și îmbunătățirea lor. Ne ocupăm de potențialul pe care metodele analitice bazate pe reguli aleatorii îl au, în scopul îmbunătățirii metodei de estimare și a analizei folosite în mod tradițional până acum, în baza unor tehnici statistice (parametrice) clasice. În acest sens, regulile aleatorii pot fi o alternativă plauzibilă sau complementară la rezultatele obținute prin folosirea tehnicilor modelării ecuației structurale (SEM), cele folosite, în general, în ultimele decenii pentru estimarea modelelor complexe ale comportamentului consumatorilor.

Această lucrare prezintă o metodologie a „descoperirii cunoașterii în baza de date” (KDD), structurată *ad hoc*, care să poată fi aplicată în modelarea (cauzală) din marketing.

## 1. Background

Marketing academics and practitioners have pointed out the necessity for knowing and explaining the consumer's behaviour patterns in a manner increasingly efficient. Firms focused on final markets are immersed in highly competitive systems in which it is needed that their decision processes to be as correct as possible.

In this regard, models of consumer behaviour, inasmuch as they are marketing models, are considered as a specific case of Marketing Management Support System (MkMSS), and throughout the time have demonstrated to be a source of transcendental relevance for the development of marketing science (van Bruggen & Wierenga, 2000).

Notwithstanding, current models of consumer behaviour do not seem to cover all the necessities that it should supposedly satisfy a model which aims to aid on the marketing decision making. With respect to this, based on Gatignon (2000), future models, considering both their theoretical and technical aspects, which try to explain consumers' decision making will have to be clearly focused on users' (demand side) requirements of such models. That is to say, models must be more complete, flexible, and customized to the strategic singularities of the competitive environment which their users operate in. Thus, as the main problem that actually face firms oriented to consumer markets is not the availability of information (data), but the possession of necessary level of knowledge to take the right decisions, the use of avant-garde behavioural models able to exploit it may represent an essential source of competitive advantage.

Doubtless, it is time to rethink the role of marketing research and modelling, in order to develop more adequate analytical methods to tackle the current business environments and decisional scenarios (Wind, 2006). It is expected that MkMSS will tend to improve their performance taking advantage of synergies caused by the integration of modelling estimation techniques based on classic econometric with expert systems based on artificial intelligence.

Specifically, considering the three pillars in which marketing modelling is based (Roberts, 2000), and more specifically the consumer behaviour modelling, we focus our paper on one of them, i.e.: the modelling estimation techniques and its improvement. We treat the potentials that analytical methods based on fuzzy rules have to evolve the method of estimation and analysis traditionally used till now, basically based on classic statistical (parametric) techniques. In this sense, fuzzy rules can be a plausible alternative or complement to the results obtained by using Structural Equation Modelling (SEM) techniques which have been the ones usually used in the last decades to estimate complex models of consumer behaviour.

This paper presents a Knowledge Discovery in Database (KDD) methodology developed *ad hoc* to be applied in marketing (causal) modeling. A *descriptive rule*

Se aplică metoda descriptivă a inducției regulii (Lavrac et al., 2004) pentru stabilirea regulilor individuale care determină tiparele informaționale care prezintă un interes deosebit. În acest scop, ne bazăm pe regulile asocierii aleatorii, nu înainte însă de a stabili variabilele de antecedentă și consecință: adică, folosim un model teoretic (cauzal) de referință, în scopul supervizării procesului de învățare. Extragerea datelor se face prin sisteme genetice aleatorii, un program hibrid de calcul. Se prezintă și rezultatele experimentale ale modului său de funcționare.

## 2. Descoperirea cunoașterii pe baza regulilor aleatorii

În linii mari, descoperirea cunoașterii în bazele de date (KDD) este un domeniu de cercetare mai recent, legat de inteligența artificială, al cărui scop principal este acela de identificare a unor tipare de date noi, potențial utile și ușor de înțeles (Fayyad, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996). În plus, KDD implică și dezvoltarea unui proces alcătuit din mai multe etape. În acest sens, căutarea de date, considerată elementul nucleu al procesului KDD, se caracterizează prin aplicarea metodelor de învățare asistată de calculator, pentru extragerea automată sau semi-automată a tiparelor sau modelelor din datele deținute (Witten & Frank, 1999).

În zilele noastre, unul dintre instrumentele cele mai utile pentru dezvoltarea unor modele descriptive este modelarea aleatorie (Lindskog, 1997), această abordare fiind utilizată pentru modelarea unui sistem cu ajutorul unui limbaj descriptiv, bazat pe logica aleatorie cu predicate aleatorii (Sugeno & Yasukawa, 1993). Predicatele aleatorii sunt redată prin intermediul regulilor DACĂ...ATUNCI, cu următoarea structură:

DACĂ  $X_1$  este  $A_1$  și ... iar  $X_n$  este  $A_n$  ATUNCI  $Y_1$  este  $B_1$  iar ...  $Y_m$  este  $B_m$

Aceste reguli stabilesc relațiile logice dintre variabilele unui sistem folosind valori calitative. O astfel de reprezentare poate fi ușor înțeleasă de om. Așadar, atât în etapa de analiză cât și în cea de interpretare a procesului de modelare, performanța se îmbunătățește, deoarece comportamentul adevărat al sistemului este dezvăluit mai eficient. De notat, totuși, că deși rațiunea umană poate înțelege fără dificultate termeni ca *înalt* sau *crește rapid*, abordarea acestor concepte printr-un proces automat este mai complexă.

Pentru utilizarea corectă a acestor evaluări calitative se folosesc variabile lingvistice (Zadeh, 1975) ce se bazează atât pe teoria multimilor aleatorii, cât și pe logica aleatorie (Zadeh, 1965), astfel că regula exemplificată mai sus este o *regulă aleatorie*. Folosirea logicii aleatorii are unele beneficii: un grad mai mare de generalitate, forță expresivă, capacitatea de a modela probleme reale și, la fel de important, o metodologie capabilă să trateze imprecizia cu toleranță. De exemplu, putem lua în considerare variabila lingvistică *vârstă*, cu valori lingvistice de tipul *adolescent*, *tânăr*, *adult* și *bătrân*.

*induction* method (Lavrac et al., 2004) is posed to discover individual rules which show information patterns of especial interest in the data. To do this, we consider fuzzy association rules, but previously setting antecedents' and consequents' variables; i.e. we use a theoretic (causal) model of reference, which is used to supervise the machine learning process. Extraction is realized by genetic fuzzy systems, a soft computing hybridization. An empirical illustration of how it works is also provided.

## 2. Knowledge Discovery Based on Fuzzy Rules

In general terms, knowledge discovery in databases (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, Piatetsky-Shapiro, Smyth & Uthurusamy, 1996). Furthermore, KDD implies the development of a process which is compounded by several stages. In this sense, data mining, which is considered as the core of KDD process, is characterized by the application of machine learning methods to automatically or semi-automatically extract patterns or models from data (Witten & Frank, 1999).

Nowadays, one of the most successful tools to develop descriptive models is fuzzy modelling (Lindskog, 1997), which is an approach used to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates (Sugeno & Yasukawa, 1993). The way to express fuzzy predicates is by means of IF-THEN rules with the following structure:

$$\text{IF } X_1 \text{ is } A_1 \text{ and } \dots \text{ and } X_n \text{ is } A_n \text{ THEN } Y_1 \text{ is } B_1 \text{ and } \dots Y_m \text{ is } B_m$$

These rules set logical relationships among variables of a system by using qualitative values. Such representation mode has the power to be easily understandable by human being. Hence, the performance of both, analysis and interpretation steps of the modelling process, improve thanks to the true behaviour of system is more effectively revealed. Notwithstanding, it should be noted that though human reasoning may manage without strain with 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 this kind of qualitative valuations, linguistic variables (Zadeh, 1975) based on both Fuzzy Sets Theory and Fuzzy Logic (Zadeh, 1965) are used, so previous exemplified rule is known as a *fuzzy rule*. The use of fuzzy logic provides several benefits as: a higher generality, expressive power, ability to model real problems and, at last but not least, a methodology to exploit tolerance in the face of imprecision. In example, we can consider the linguistic variable *age*, which linguistic values could be *teenager*, *young*, *adult*, and *old*.

Regulile aleatorii pot fi privite și ca un instrument de extragere a cunoașterii menit să descopere relațiile intrinsece dintr-o bază de date (Freitas, 2002). Astfel, cu ajutorul regulilor aleatorii se pot reprezenta relațiile existente între diferite variabile, deducându-se astfel tiparele conținute în datele studiate. În descoperirea cunoașterii, procesul prin care se deduc aceste tipare trebuie să fie automat, sau semi-automat, tiparele descoperite trebuie să fie clare și să cuprindă informații utile, iar datele trebuie să fie întotdeauna prezentate în cantități substanțiale (Witten & Frank, 2000).

Tiparele utile ne permit să emitem predicții juste despre noile date. Există două modalități extreme de exprimare a datelor: sub forma unor cutii negre, cu un comportament interior de neînțeles, sau sub forma unor cutii albe, a căror construcție lasă să se vadă structura lor interioară. Diferența constă în reprezentarea tiparelor generate într-un mod mai ușor de studiat, care poate fi folosit pentru raționamente și decizii ulterioare. Cu alte cuvinte, atunci când tiparele sunt structurate într-un mod ușor de înțeles, ele pot ajuta la explicarea datelor. Această problemă a KDD – relația dintre caracterul interpretabil și gradul de acuratețe – este la fel de importantă și în cazul modelelor aleatorii (Casillas et al., 2003a, 2003b) și va fi verificată în acest studiu.

Folosirea regulilor aleatorii în desfășurarea procesului de descoperire a cunoașterii are anumite avantaje : ele permit folosirea datelor nesigure ; iau în considerare relațiile cu variabile multiple ; rezultatele sunt ușor de înțeles de către mintea umană ; un expert poate adăuga informații suplimentare ; gradul de corectitudine poate fi ușor adaptat la necesitățile curente, iar procesul poate fi automatizat astfel încât intervenția umană să fie redusă la minimum.

Așadar, vom folosi logica aleatorie ca instrument de structurare a informației referitoare la modelul comportamental al consumatorului, într-un mod clar, ușor de înțeles și asemănător funcționării minții umane. Sistemul aleatoriu ne va permite să reprezentăm, în mod corespunzător, interdependența dintre variabile și relațiile non-lineare ce pot apărea între acestea. În sfârșit, algoritmii de reprezentare (un algoritm genetic în lucrarea de față) vor contura regulile aleatorii corespunzătoare criteriilor de interpretabilitate și exactitate impuse de experți.

Secțiunea care urmează prezintă metodologia prin care extragerea datelor cu ajutorul regulilor aleatorii se aplică la modelarea comportamentului consumatorului.

### **3. Metodologia de modelare a comportamentului consumatorului prin sisteme aleatorii genetice**

#### **3.1. Colectarea datelor**

Primul pas este acela de a colecta datele legate de variabilele care definesc modelul teoretic de referință al comportamentului consumatorului. În acest sens, așa cum se face, prin tradiție, în marketing, datele se obțin cu ajutorul unor chestionare. Așadar, în primul rând trebuie verificat modul în care cei care structurează modelele de comportament ale consumatorului tratează și realizează măsurarea variabilelor conținute

Fuzzy rules can be considered as a knowledge extraction tool 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 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 & Frank, 2000).

Useful patterns allow us to do non trivial predictions about new data. There are two extremes to express a pattern: like black boxes, whose internal behaviour is incomprehensible; and like white boxes, whose construction reveals the pattern structure. The difference lies in whether the generated patterns are represented with 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 trouble of 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 multi-variable 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; 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, legible, and close to the human being way. The fuzzy system will allow us to properly represent the interdependence of variables and the non-linear relationships that could exist among them. Finally, optimization algorithms (a genetic algorithm in this paper) will design the fuzzy rules to meet the interpretability and accuracy criteria imposed by the expert.

The following section introduces the methodology followed for applying data mining by means of fuzzy rules to consumer behaviour modelling.

### **3. A Methodology for Consumer Behavior Modelling by Genetic Fuzzy Systems**

#### **3.1. Data Gathering**

First step is to collect the data related to the variables defining the theoretic consumer behavior model of reference. In this sense, as it has been traditionally done in marketing, data are obtained by means of a questionnaire. Thus, firstly, attention should be paid to how consumer behavior modelers face and develop the measurement process of variables that complex behavioral models contain; i.e. usually, latent/unobserved variables. Its understanding is necessary in order to adequately approach the starting



în modelele complexe de comportament al consumatorului ; acestea sunt, de obicei, variabile latente/neobservate. Înțelegerea lor este necesară pentru abordarea corectă a punctului de început al KDD, pentru obținerea soluțiilor potrivite, corespunzătoare pentru datele specifice constatate în modelarea comportamentului consumatorului.

Se poate spune că măsurarea tendințelor acestor variabile latente în cadrul modelării din marketing poate fi clasificată în două grupe, în funcție de capacitatea acestor construcții de a fi sau nu măsurabile prin intermediul unor variabile observate (indicatori) ; ne referim la existența sau inexistența unei corespondențe unu-la-unu între construcție și măsurătorile acesteia. Desigur, deși autorii modelelor de comportament al consumatorilor par să pornească de la o *filosofie a interpretării operaționale*, ar fi mai bine ca ulterior să se recurgă la o *filosofie de interpretare parțială*, care să facă diferența între variabilele neevidente (construcții) și cele evidente (indicatori). Această ultimă abordare a măsurării, predominantă în disciplina modelării din marketing, ia în considerare mai mulți indicatori ai construcției subordonate în același timp – ei sunt imperfecti, dacă sunt luați separat, dar demni de încredere, luați împreună – pentru obținerea unor unități de măsură valabile (Steenkamp și Baumgartner, 2000). Așadar, vom lua în considerare măsurătorile în etapa de prelucrare a datelor.

### 3.2. Prelucrarea datelor

Etapă următoare constă în adaptarea datelor colectate la o schemă definită prin metodele de învățare prin reguli aleatorii. Așadar, abordarea noastră metodologică trebuie să țină seama de trăsăturile speciale ale datelor deținute (cu ajutorul câtorva elemente sau indicatori care să descrie o variabilă specifică) atunci când se începe procesul de adaptare a variabilelor observate la metoda de învățare prin reguli aleatorii. Abordarea intuitivă poate reduce direct elementele anumitor variabile la o singură valoare (de exemplu, prin media aritmetică). O altă posibilitate este aceea de a extinde valabilitatea unui exemplu cu elemente multiple (rezultatul unui chestionar completat de un consumator) la mai multe exemple ale unui singur element și, în consecință, de a reduce cantitatea de date printr-un sistem de selecție (Casillas, Martínez-López & Martínez, 2004).

Problema acestor abordări o reprezintă faptul că informațiile respective trebuie transformate, astfel că există riscul pierderii de informații relevante. Noi propunem un sistem mai complex, care permite lucrul cu formatul original, fără etape anterioare de prelucrare : aplicarea *transformării aleatorii cu elemente multiple*. Astfel, un operator *T-conorm* (de exemplu maximum), folosit în mod tradițional în logica aleatorie pentru cumularea informațiilor parțiale date de fiecare element în timpul procesului de deducție. Nefiind vorba de o prelucrare premergătoare a datelor, ci de o componentă a proiectului de învățare automatizată, detaliile de abordare a elementelor sunt descrise în Secțiunea 3.4.2.

point of the KDD process, so to give suitable and adapted solutions to the specific data we find in consumer behavior modeling.

It can be said that measuring streams for these latent variables in marketing modeling can be classified into two groups depending on if they state that these constructs can or cannot be perfectly measured by means of observed variables (indicators); i.e., the existence or not of a one-to-one correspondence between a construct and its measurement. Certainly, though consumer behavior modelers tended to make use in the beginning of what was known as the *operational definition philosophy*, a more convenient and reasonable position is that ulteriorly 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 modeling discipline, poses to jointly consider multiple indicators – imperfect when considered individually, though reliable when considered altogether – of the subjacent construct to obtain valid measures (Steenkamp & Baumgartner, 2000). Hence, we will take this measurement approach into account when facing how to process the data.

### **3.2 Data Processing**

Next, it is necessary to adapt the collected data to a scheme easily tractable by fuzzy rule learning methods. Therefore, our methodological approach should be aware of the special features of the available data (with several items or indicators to describe a specific variable) when adapting the observed variables to a fuzzy rule learning method. An intuitive approach could directly reduce the items of certain variables to a single value (e.g., by arithmetic mean). Another possibility would be to expand any multi-item example (the result of a questionnaire filled out by a consumer) to several single-item examples and, subsequently, reduce the data size with some instance of selection process (Casillas, Martínez-López & Martínez, 2004).

The problem of these approaches is that the data must be transformed, so relevant information may be lost. We propose a more sophisticated process that allows working 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, is applied to aggregate the partial information given by each item during the inference process. Since it is not pre-processing data but a component of the machine learning design, the details of that treatment of the items is described in Section 3.4.2.

### **3.3. Reprezentarea și includerea cunoașterii provenite de la experți**

Anumite probleme trebuie rezolvate în această etapă : setul de variabile care urmează să fie modelat, transformarea scalelor de marketing folosite pentru măsurarea acestor variabile, în structuri semantice aleatorii și reguli aleatorii (relații între construcții). Propunem câteva variante de calcul pentru aceste componente. Toate se bazează pe capacitatea expertului în marketing de a-și exprima cunoștințele într-un format uman inteligibil, prin intermediul logicii aleatorii.

#### **3.3.1. Semantică aleatorie derivată din cunoașterea expertului**

După ce operatorul care realizează modelul stabilește atât construcțiile teoretice cât și variabilele evidente asociate cu fiecare dintre acestea (adică modelul de măsurare), scala originală de marketing trebuie transformată pentru măsurarea acelor variabile evidente, în termeni lingvistici. În acest punct se pot folosi mai multe tipuri de scale de măsurare. Pentru a simplifica această problemă, lucrarea de față folosește scala de tip Lickert, scale semantice diferențiale și de evaluare – cele mai des folosite pentru aceste modele. Transformarea se face ținând cont de următoarele trei probleme :

1. Stabilirea *numărului termenilor lingvistici folosiți pentru fiecare variabilă*. Un număr impar este potrivit, deoarece, în cazul nostru, este importantă exprimarea lingvistică a unui concept « mediu » sau « indiferent ». Cum scalele de interval tradiționale folosite în marketing includ între 5-9 grade (puncte de scală) diferite, folosirea unui număr de trei sau patru termeni lingvistici (seturi aleatorii) este suficientă pentru a realiza harta acestor valori.

2. Stabilirea *funcției de membru*, care definește comportamentul anumitor variabile aleatorii. În acest sens, comportamentul ce poate fi calculat, în mare, prin folosirea unor funcții de membru lineare (trapezoidale sau triunghiulare) în comparație cu cele non-lineare (Gaussiene), pentru caracterizarea seturilor aleatorii. Ca atare, suntem de părere că folosirea funcțiilor lineare este preferabilă, deoarece facilitează interpretarea ulterioară a relațiilor.

3. Stabilirea formelor funcției de membru. Din acest punct de vedere, propunem impunerea unor proprietăți în scopul unei interpretări optime. Valorile extreme ale intervalului ar trebui să aibă un grad de membru de 1 la extreme. Valoarea medie a intervalului ar trebui să aibă gradul de membru 1 în zona medie. În mod similar, vom folosi semantica aleatorie definită de Ruspini (suma gradelor de membru ale fiecărei valori pentru stabilirea setului de termeni lingvistici este egală cu 1) pentru ca interpretarea să fie corectă. În sfârșit, pentru clarificarea statistică a semnificației fiecărui termen lingvistic, propunem același grad de acoperire. Astfel, se definesc formele funcției de membru, unde, pentru un set  $S=\{\min,\dots,\max\}$  care definește intervalul, se susține următoarea condiție:

### 3.3. Representation and Inclusion of Expert Knowledge

Several issues should be tackled at this step: the set of variables to be modeled, the transformation of marketing scales used for measuring such variables into fuzzy semantic and the fuzzy rule structure (relations among constructs). We suggest some approaches to fix these components. All of them are based on the marketing expert's capability to express his knowledge in a humanly understandable format by fuzzy logic.

#### 3.3.1. Fuzzy Semantics from Expert Knowledge

Once the marketing modeler has finally determined both, the theoretical constructs and the observed variables associated with each one (i.e. the measurement model), a transformation of the original marketing scales used for measuring those observed variables into linguistic terms should be done. At this point, several marketing scale types can be used for its measurement. With the aim of simplifying the problem, in this paper we focus on Likert-type, differential semantic and rating scales, which are the most commonly used in these models. The transformation should be practiced taking into account three main questions:

1. The *number of linguistic terms* to be used for each variable must be defined. An odd number seems to be a good approach since in our case it is useful to linguistically express the "medium" or "unconcerned" concept. Since traditional interval scales used in marketing usually present between 5 to 9 different degrees (i.e. points of the scale), the use of three or five linguistic terms (fuzzy sets) is enough to map these values.

2. The *membership function type* defining the behavior of certain fuzzy variables should be also defined. In this sense, such behavior can be broadly treated considering the use of linear (trapezoidal or triangular) vs. non linear (Gaussian) membership functions to characterize the fuzzy sets. In this respect, we pose that it is more appropriate to use linear functions, inasmuch as it facilitates the latter interpretation of relations.

3. The *membership function shapes* should also be fixed. In this respect, we propose to impose some properties in order to ensure good interpretability. Extreme values of the interval should have a membership degree 1 to extreme labels. Mean value of the interval should have membership 1 to medium label. Likewise, we consider strong Ruspini's fuzzy semantics (the sum of the membership degrees of every value to the set of linguistic terms is 1) in order to ensure good interpretability. Finally, in order to statistically unbiased the significance of every linguistic term, we impose the same covering degree. Thus, we define the membership function shapes where, given the set  $S=\{\min,\dots,\max\}$  defining the interval, they hold the following condition:

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

unde  $l$  este numărul de termeni lingvistici, iar  $A = \{A_1, \dots, A_l\}$  este setul acestora.

Astfel, Figura 1 prezintă un exemplu bazat pe transformarea unei scale de evaluare cu nouă puncte (o scală tipică folosită în marketing pentru măsurarea variabilelor/indicatorilor observați cu privire la o anumită construcție) într-o semantică aleatorie cu trei termeni lingvistici: *nivel scăzut, mediu și înalt*.

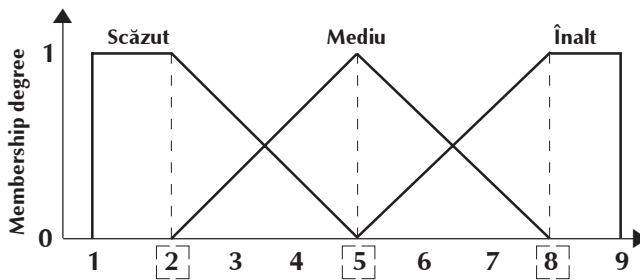


Figura 1. Semantica aleatorie rezultată din transformarea unei scale de marketing cu 9 puncte (scală de evaluare)

### 3.3.2. Variabile lingvistice incluse/rezultate din cunoașterea de către expert

Odată ce structura modelului a fost stabilită de către expertul în marketing pe baza modelului teoretic, se folosesc regulile aleatorii pentru a stabili relația dintre variabilele elementelor introduse (antecedente) și cele rezultate (consecințe). Desigur, ipotezele conținute în model pot fi folosite direct pentru definirea structurilor DACĂ...ATUNCI pe baza dependenței dintre variabile. Astfel se obține baza pentru regula aleatorie, pentru fiecare consecință (construcție endogenă) luată în calcul și pentru seturile corespunzătoare de antecedente.

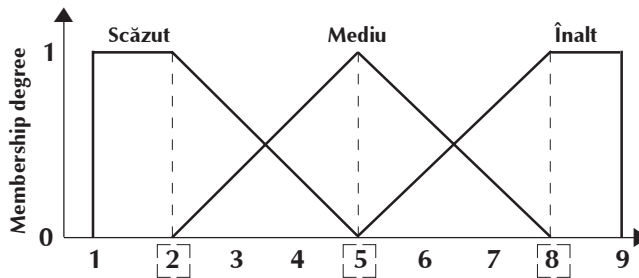
De exemplu, dacă, în scop ilustrativ, se ia modelul asociat cu teoria acțiunii raționale (Ajzen & Fishbein, 1980, structura regulii aleatorii, care reprezintă relațiile binecunoscute între elementele "atitudine" și "normă subiectivă" cu consecința "intenție", vor avea următoarea formă:

**DACĂ** atitudinea este  $A_1$  și norma subiectivă este  $A_2$ , **ATUNCI** intenția este B.

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

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

To sum up, Figure 1 shows an example based on the transformation of a nine-point rating scale (a typical marketing scale used to measure the observed variables/indicators related to certain construct) into a fuzzy semantic with the three linguistic terms *Low*, *Medium*, and *High*.



**Fig. 1. Fuzzy semantic from a transformation of a 9-point marketing scale (rating scale)**

### 3.3.2. Input/Output Linguistic Variables from Expert Knowledge

Furthermore, once the structure of the model has been fixed by the marketing expert under the base of the theoretic model, fuzzy rules are used to relate input (antecedents) with output (consequents) variables. Obviously, hypotheses contained in the model can be directly used to define IF-THEN structures by considering the dependencies shown among the variables. Thus, we obtain a fuzzy rule base for each consequent (endogenous construct) considered and its respective set of antecedents.

For example, if we took for illustrative purposes the model associated with the Theory of Reasoned Action (Ajzen & Fishbein, 1980), the fuzzy rule structure which represents the widely known relations between the elements “attitude” and “subjective norm” with the consequent “intention” will have the following form:

**IF** Attitude is  $A_1$  and SubjectiveNorm is  $A_2$  **THEN** Intention is B.

### 3.4. Procesul de culegere a datelor

După fixarea variabilelor lingvistice care redau corect informația în cauză, se utilizează un proces automat de învățare, pentru extragerea automată a cunoștințelor existente în baza de date. Fără îndoială, acest proces este cel mai important din punctul de vedere al KDD.

Așa cum s-a arătat în secțiunea Generalități, în această lucrare ne interesează inducția descriptivă. Așadar vom folosi stilul GA-Michigan pentru deducerea unor reguli relevante la nivel individual. Se iau în considerare două criterii de calitate – suport (măsura în care regula este reprezentativă pentru setul de date) și încredere (cât de exactă este relația definită de regulă). Intuitiv se poate verifica dacă un suport mai puternic duce la o mai mare dificultate în menținerea unui grad înalt de încredere. Pentru studiul unitar al celor două criterii, propunem folosirea GA cu obiective multiple, deoarece în acest fel se obțin rezultate bune atunci când obiectivele sunt multiple și contradictorii. Secțiunea următoare descrie principalele elemente ale acestei metode pe care o propunem.

#### 3.4.1. Structura bazată pe regula aleatorie

La extragerea datelor este esențială folosirea unui proces de învățare cu un grad înalt de interpretabilitate. În acest scop am optat pentru o descriere compactă, bazată pe forma normală disjunctă (DNF). Acest tip de structură cu reguli aleatorii are următoarea formă:

DACĂ  $X_1$  este  $\tilde{A}_1$  și ... iar  $X_n$  este  $\tilde{A}_n$  ATUNCI  $Y_1$  este  $B$

Unde fiecare variabilă antecedentă  $X_i, i \in \{1, \dots, n\}$  are valoarea de forma unui set de termeni lingvistici  $\tilde{A}_i = \{A_{i1} \text{ sau } \dots \text{ sau } A_{in}\}$ , ai cărui membri sunt alăturați de un operator disjunctive. Suma delimitată  $\min\{1, a+b\}$  este folosită ca *T-conorm*. Această structură este un support natural care permite absența unor variabile antecedente în fiecare regulă, doar prin transformarea lui  $\tilde{A}_j$  într-un set complet de termeni lingvistici disponibili.

#### 3.4.2. Transformări aleatorii cu elemente multiple

Pentru un studiu corect al setului de indicatori disponibili pentru fiecare variabilă antecedentă și consecutivă (așa cum s-a arătat în Secțiunea 3.2), propunem o extindere a calculului gradului de apartenență, așa numita *transformare aleatorie cu elemente multiple*. Procesul în cauză se bazează pe o grupare a informațiilor parțiale date de fiecare element în parte. Se dau  $X_i$  și  $Y_j$  măsurate de vectorii elementelor  $\vec{x}_i = (x_1^{(i)}, \dots, x_{h_i}^{(i)}, \dots, x_{p_i}^{(i)})$  și respectiv  $\vec{y} = (y_1, \dots, y_{l'}, \dots, y_q)$ , iar propozițiile aleatorii  $X_i$  este  $\tilde{A}_i$  și  $Y$  este  $B$  sunt interpretate astfel:

### 3.4. Data Mining Process

Once the linguistic variables that properly represent the tackled information have been fixed, a machine learning process must be used to automatically extract the knowledge existing in the database. This process is, without any doubt, the most important issue from the KDD point of view.

As mentioned in the Background Section, in this paper we are interested in descriptive induction. Therefore, we will use GAs Michigan-style to obtain rules individually relevant. We consider two quality criteria, support (degree of representativity of the rule with respect to the set of data) and confidence (degree of accuracy of the relation shown by the rule). It is intuitive to check that the higher the support, the higher the difficulty to maintain high degrees of confidence. To jointly consider both criteria, we propose the use of *multiobjective GAs*, as they offer good results when working with multiple contradictory objectives. The next section describes the main elements of this method we propose.

#### 3.4.1. Fuzzy Rule Structure

In data mining it is crucial to use a learning process with a high degree of interpretability. To do that, we opt for a compact description based on the disjunctive normal form (DNF). This kind of fuzzy rule structure has the following form:

IF  $X_1$  is  $\tilde{A}_1$  and ... and  $X_n$  is  $\tilde{A}_n$  THEN  $Y_1$  is B

where each input variable  $X_i$ ,  $i \in \{1, \dots, n\}$  takes as a value a set of linguistic terms  $\tilde{A}_i = \{A_{i_1} \text{ sau } \dots \text{ sau } A_{i_m}\}$ , whose members are joined by a disjunctive operator. We use the bounded sum  $\min\{1, a+b\}$  as *T-conorm*. The structure is a natural support to allow the absence of some input variables in each rule, simply making  $\tilde{A}_i$  to be the whole set of linguistic terms available.

#### 3.4.2. Multi-item Fuzzification

In order to properly consider the set of indicators 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  $\vec{x}_i = (x_1^{(i)}, \dots, x_{h_i}^{(i)}, \dots, x_{p_i}^{(i)})$  and  $\vec{y} = (y_1, \dots, y_r, \dots, y_q)$ , respectively, the fuzzy propositions  $X_i$  is  $\tilde{A}_i$  and  $Y$  is B are respectively interpreted as follows:



$$\mu_{\vec{A}_i}(\vec{x}_i) = \min \left\{ 1, \bigcup_{h_i=1}^{p_i} \sum_{A \in \vec{A}_i} \mu_A(x_{h_i}^{(i)}) \right\}$$

$$\mu_{\vec{B}}(\vec{y}) = \bigcup_{t=1}^q \mu_{B_t}(y_t),$$

unde  $\bigcup$  este T-conorm (maximum în această lucrare).

### 3.4.3. Cercetare pe subgrupe

Pentru realizarea procesului de inducție a regulilor descriptive, s-a aplicat o metodă având anumite similarități cu tehnica cercetării pe subgrupe – folosită pe larg la clasificarea regulilor de învățare (Lavrač, 2004) – prin care proprietatea de interes este clasa asociată cu variabilele consecutive. Așadar, am încercat să grupăm seturile de date în subgrupe diferențiate, incluzând în fiecare dintre acestea acele exemple care sunt reprezentate de consecutivă, cu scopul de a defini un set reprezentativ de reguli pentru fiecare subgrupă. Practic, abordarea cea mai comună se bazează pe aplicarea algoritmului desemnat pentru fiecare subgrupă de date, care satisface setul de proprietăți ale consecutivei.

Totuși, în locul acestei abordări, propunem un alt algoritm de cercetare simultană a subgrupelor. Această variantă permite formarea unor nișe de reguli aleatorii diferențiate prin consecutivă; acestea sunt optimizate în paralel, pentru a genera în final un set de soluții suboptimale pentru fiecare clasă de consecvente. În scopul derulării acestui proces simultan, așa cum se arată în următoarele secțiuni, am variat conceptul de dominanță cu obiectiv multiplu, făcând ca operatorii genetici să acționeze doar asupra antecedentelor regulilor.

### 3.4.4. Schemă de codificare

Fiecare individ dintr-o mulțime reprezintă o mulțime aleatorie, adică un algoritm genetic tip Michigan. Schema de codificare trebuie să fie binară, pentru a reprezenta antecedenta, și integrală pentru consecutivă. Astfel, alela "1" din antecedentă înseamnă că termenul lingvistic legat de genă este folosit la variabila corespunzătoare. Pentru consecutivă se codifică direct indicele termenului lingvistic folosit. Astfel, mărimea de codificare a unei reguli aleatorii a DNF este egală cu suma numărului de termeni lingvistici folosiți pentru fiecare variabilă a antecedentei, plus numărul de variabile ale consecutivei. De exemplu, dacă am avea trei termeni lingvistici pentru fiecare variabilă, regula [DACĂ X1 este Mic și X2 este {Mediu sau Mare} ATUNCI Y este Mediu] ar putea fi codificată ca [100 01112].

$$\mu_{\vec{A}_i}(\vec{x}_i) = \min \left\{ 1, \bigcup_{h_i=1}^{P_i} \sum_{A \in \vec{A}_i} \mu_A(x_{h_i}^{(i)}) \right\}$$

$$\mu_{\vec{B}}(\vec{y}) = \bigcup_{t=1}^q \mu_B(y_t),$$

with  $\bigcup$  being a T-conorm (the maximum in this paper).

### 3.4.3. Subgroup Discovery

To do the descriptive rules induction process, we have applied a method with certain similarities to the subgroups discovery technique –widely used in classification learning rules (Lavrač, 2004)–, where the property of interest is the class associated with the variables of the consequent. Therefore, we try to group the set of data into differentiated subgroups, including in each of them those examples represented by the consequent with the aim of discovering a representative set of rules for each subgroup. In this regard, the most usual approach is based on running the algorithm designed for each subgroup of data which satisfies the property set for the consequent.

However, instead of this approach, we carry out a simultaneous subgroup discovery in the algorithm we propose. This variant allows us to form niches of fuzzy rules differentiated by the consequent which are optimized in parallel to finally generate a set of suboptimal solutions for each class of the consequent. With the aim of developing this simultaneous process, as it is shown in the next sections, we vary the concept of multiobjective dominance by making the genetic operators act only on the antecedents of the rules.

### 3.4.4. Coding Scheme

Each individual of the population represents a fuzzy rule; i.e. a Michigan-style genetic algorithm. The coding scheme will be binary to represent the antecedent and whole for the consequent. Thus, the allele “1” in the antecedent part means that the linguistic term related to the gene is used in the corresponding variable. For the consequent, we will directly code the index of the linguistic term used. Hence, the size to code a DNF fuzzy rule is equal to the sum of the number of linguistic terms employed in each input variable (antecedent) plus the number of output variables. For instance, if we had three linguistic terms for each variable, the rule [IF X1 is Small and X2 is {Medium or High} THEN Y is Medium], would be coded as [100 01112].

### 3.4.5. Funcții obiective

În acest algoritm, se iau în considerare două criterii cel mai frecvent folosite pentru evaluarea calității regulilor de asociere (Dubois, Prade & Sudkamp, 2005): suport și încredere. Totuși, am adaptat calculul acestor criterii la regulile de asociere aleatorii, ținând cont și de caracteristicile speciale ale variabilelor cu elemente multiple (elemente ale modelului) cu care am lucrat.

**SUPPORT.** Această funcție obiectivă evaluează gradul de reprezentare a unei reguli aleatorii în setul de date analizat. Ea se calculează ca grad mediu sub incidența regulii, ținând cont de fiecare dintre aceste date (răspunsuri la nivel individual). Pentru a obține gradul de acoperire se iau în considerare gradele de apartenență, grupate, în relație cu diferite variabile, adică atât cu setul de antecedente, cât și cu setul de consecvente. Unitatea de măsură a suportului (pentru maximizare) pentru o regulă aleatorie R se definește astfel:

$$Support(R) = \frac{1}{N} \sum_{e=1}^N T(\mu_A(x^{(e)}), \mu_B(\vec{y}^{(e)})),$$

unde N este mărimea bazei de date (mărimea eșantionului sau numărul de respondenți),  $x^{(e)} = (\vec{x}_1^{(e)}, \dots, \vec{x}_n^{(e)})$  și  $\vec{y}^{(e)}$  reprezintă incidența cu elemente multiple e a antecedentei și respectiv a consecutivei, T este produsul T-norm, iar  $\mu_A(x^{(e)}) = \min_{i \in \{1, \dots, n\}} \mu_{\vec{A}_i}(\vec{x}_i^{(e)})$  gradul de acoperire an antecedentei din regula R pentru acest exemplu (adică se consideră T-norm al valorii minime pentru interpretarea conectorului « și » a regulii aleatorii). În plus, trebuie relevat și faptul că am folosit sistemul aleatoriu cu elemente multiple prezentat în secțiunea 3.4.2 pentru calcularea lui  $\mu_{\vec{A}_i}(\vec{x}_i^{(e)})$  și  $\mu_B(\vec{y}^{(e)})$ .

**ÎNCREDERE.** Această funcție obiectivă măsoară gradul de încredere al relației între antecedentă și consecutivă, descris de regula aleatorie descrisă mai sus. Am utilizat un grad de încredere care evită acumularea unor elemente cardinale joase [4]. Se calculează (pentru maximizare) după cum urmează:

$$Confidence(R) = \frac{\sum_{e=1}^N T(\mu_A(x^{(e)}), I(\mu_A(x^{(e)}), \mu_B(\vec{y}^{(e)})))}{\sum_{e=1}^N \mu_A(x^{(e)})},$$

Se utilizează implicația S aparținând lui Dienes  $I(a,b) = \max\{1 - a, b\}$ . Se ia din nou în considerare valoarea T-norm a produsului și sistemul aleatoriu cu elemente multiple.

### 3.4.5. Objective Functions

In this algorithm, we consider the two criteria most frequently used to value the quality of the association rules (Dubois, Prade & Sudkamp, 2005): support and confidence. However, we adapt the calculus of these criteria to fuzzy association rules, also considering the especial characteristics of the multi-item variables (elements of the model) which we work with.

**Support.** This objective function values the degree of representation of certain fuzzy rule on the set of data analyzed. It is calculated as the average degree covered by the rule considering every one of these data (individuals' responses). To obtain the degree of cover we conjointly consider the membership degrees in relation to the diverse variables; i.e. the set of antecedents as well as the consequent. The measure of support (for maximization) for a fuzzy rule R comes defined as follows:

$$Support(R) = \frac{1}{N} \sum_{e=1}^N T(\mu_A(x^{(e)}), \mu_B(\vec{y}^{(e)})),$$

where N is the size of the database (the sample size or number of respondents),  $x^{(e)} = (\vec{x}_1^{(e)}, \dots, \vec{x}_n^{(e)})$  and  $\vec{y}^{(e)}$  is the eth instance multi-item of input and output respectively, T the *product* T-norm, and  $\mu_A(x^{(e)}) = \min_{i \in \{1, \dots, n\}} \mu_{\tilde{A}_i}(\vec{x}_i^{(e)})$  the coverage degree of the antecedent of the rule R for this example (i.e. it is considered the T-norm of the minimum to interpret the connector "and" of the fuzzy rule). Also, it is convenient to point out that we employ the multi-item fuzzification shown in section 3.4.2 to calculate  $\mu_{\tilde{A}_i}(\vec{x}_i^{(e)})$  and  $\mu_B(\vec{y}^{(e)})$ .

**Confidence.** This objective function measures the reliability of the relationship between antecedent and consequent described by the analyzed fuzzy rule. We have used a confidence degree that avoids accumulation of low cardinalities [4]. It is computed (for maximizing) as follows:

$$Confidence(R) = \frac{\sum_{e=1}^N T(\mu_A(x^{(e)}), I(\mu_A(x^{(e)}), \mu_B(\vec{y}^{(e)})))}{\sum_{e=1}^N \mu_A(x^{(e)})},$$

The Dienes' S-implication  $I(a,b) = \max\{1 - a, b\}$  is used. We consider again T-norm of product and multi-fuzzification.

### 3.4.6. Schema evolutivă

S-a ales o abordare generativă cu strategia de înlocuire a obiectivului multiplu NSGA-II (Deb et al., 2002). S-a folosit o selecție binară bazată pe distanța dintre elementele aglomerării, în spațiul funcțional obiectiv. Pentru deducerea corectă a rezultatului de subgrup este necesară redefinirea conceptului de dominare. Astfel, o soluție (regulă) domină o alta atunci când, pe lângă faptul că toate obiectivele sunt egale la valoare minimă, iar unul este îmbunătățit, ea prezintă aceeași consecutivă ca și cealaltă regulă. Deci, regulile cu consecutivă diferită nu se domină una pe cealaltă. Ca urmare, am impus algoritmului să formeze un număr de nișe de căutare (seturi Pareto) egal cu diferitele consecutive (subgrupe).

### 3.4.7. Operatori genetici

Mulțimea inițială este construită prin definirea unui număr de grupuri (egale ca dimensiuni) la fel de mare ca și numărul consecutivelor diferite. În fiecare dintre acestea, se generează cromozomi prin fixarea acestor consecutive și prin realizarea la întâmplare a unei construcții anterioare simple, în care fiecare variabilă inclusă este corelată cu un termen lingvistic. Cei doi operatori de reproducere acționează doar în partea anterioară a regulii. Acest fapt permite menținerea constanței dimensiunii fiecărei subgrupe. Astfel algoritmul poate explora independent, dar simultan, fiecare grup.

Am folosit un operator transversal cu puncte multiple, selectând două puncte de încrucișare (în zona anterioară); lanțul central este interschimbabil. Operatorul selectează la întâmplare o variabilă a anterioarei regulii aleatorii codificate în cromozom și desfășoară una din următoarele trei operații: *expansiune*, deplasând la 1 o genă din variabila selectată; *contractie*, deplasând la 0 o genă din variabila selectată; sau *transfer*, deplasând la 0 o genă din variabilă și la 1 gena imediat anterioară sau consecutivă acesteia. Alegerea unuia dintre aceste mecanisme se face aleatoriu în limita alegerilor posibile (de exemplu contractia nu poate fi aplicată dacă doar o singură genă a variabilei selectate are alela 1).

## 4. Ilustrare empirică a performanțelor obținute prin această metodologie

### 4.1. Comentarii anterioare despre modelul folosit pentru ilustrarea empirică

Verificarea experimentală a metodei inductive a regulii aleatorii prezentate de noi s-a făcut pe baza unui model cauzal propus și de Novak, Hoffman & Yung (2000). Acesta analizează „transfigurarea” consumatorilor în medii interactive asistate de calculator.

Pentru introducerea pe scurt a acestui concept, menită să familiarizeze cititorii cu caracteristicile acestei variabile folosite la prezentarea în detaliu a aplicației empirice a acestei metodologii, vom menționa câteva idei despre ea. Termenul de „transfigurare” a fost preluat recent din psihologia motivațională și adaptat cu succes pentru explicarea comportamentului consumatorilor pe internet (Hoffman & Novak, 1996; Korzan, 2003;

### 3.4.6. Evolutionary Scheme

A generational approach with the multi-objective NSGA-II replacement strategy (Deb et al., 2002) is adopted. A binary tournament selection is used based on the crowding distance in the objective function space. To correctly develop the simultaneous subgroup discovery we will need to redefine the concept of dominance. In order to do this, one solution (rule) will dominate another when, as well as equaling as minimum all the objectives and improving in one of them, it presents the same consequent as the other rule. Hence, those rules with different a consequent do not dominate each other. Consequently, we force the algorithm to form so many niches of search (Pareto sets) as diverse consequents (subgroups) are considered.

### 3.4.7. Genetic operators

The initial population is built defining so many groups (equal in size) as there are different consequents. In each of them, chromosomes are generated fixing such consequents and randomly building a simple antecedent where each input variable is related to a linguistic term. The two operators of reproduction only act in the part of the antecedent of the rule. This fact ensures that the size of every subgroup in the population is constant. In this way, we allow the algorithm to independently explore, but simultaneously, each group.

We employ a multipoint crossover operator which selects two crossover points (in the part of the antecedent) and interchanges the central sub-chain. The operator of mutation randomly selects a variable of the antecedent of the fuzzy rule coded in the chromosome and carries out some of the three following operations: *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 a gene of the selected variable has the allele 1)

## 4. Empirical illustration of the methodology's performance

### 4.1 Previous commentaries about the model used for the empirical illustration

The experimentation of the descriptive rule induction method we present has been made based on a causal model already proposed by Novak, Hoffman & Yung (2000). It analyzes the consumer's flow state in interactive computer-mediated environments.

In order to briefly introduce this concept, so the reader better understands the variable we want to explain in this empirical application of our methodology, we now synthetically present some ideas about it. Flow has been recently imported from motivational psychology and successfully adapted to explain consumer behavior phenomena on the Web (Hoffman & Novak, 1996; Korzan, 2003; Luna, Peracchio & De Juan, 2002; Novak, Hoffman & Duhachek, 2003; Novak, Hoffman & Yung, 2002).

Luna, Peracchio & De Juan, 2002; Novak, Hoffman & Duhachek, 2003; Novak, Hoffman & Yung, 2002). În linii mari, această stare se definește ca „procesul experienței optime”, sau starea mentală prin care trec uneori indivizii atunci când sunt adânc implicați în anumite evenimente, obiecte sau activități (Csikszentmihalyi, 1975, 1977). Conceptul a fost adaptat la mediul virtual. În acest context, starea de transfigurare este atinsă atunci când consumatorul este atât de adânc implicat în navigarea pe internet, încât „nimic altceva nu mai contează” (Hoffman & Novak, 1996: p. 57).

Deși modelul folosit pentru experiment are 12 elemente (construcții) interconectate, cu sisteme bazate pe 6 reguli aleatorii, datorită constrângerii de spațiu ne vom axa, în această lucrare, pe patru antecedente primare ale transfigurării consumatorilor. Mai exact, se consideră patru construcții (viteza de interacțiune, abilitate/comandă, provocare/entuziasm și tele-prezență/pierderea noțiunii timpului) ca antecedente pentru starea de transfigurare a consumatorilor (consecutiva). În acest sens, s-a emis ipoteza că aceste patru elemente sunt legate pozitiv de construcția centrală a acestui model.

Majoritatea elementelor construcției au fost evaluate cu scala Lickert cu 9 puncte (adică o scală metrică), cu excepția unuia, care a fost măsurat cu ajutorul unei scale ordinale. Semantica aleatorie a fost aplicată tuturor variabilelor, așa cum se arată în Figura 1.

Baza de date cuprinde 1.154 exemple (răspunsurile consumatorilor). Am aplicat algoritmul de 10 ori, obținând următoarele valori ale parametrilor: 300 generări, dimensiunea grupului 100 persoane, probabilitatea de încrucișare 0,7 și probabilitatea mutațiilor per cromozom 0,1.

#### **4.2. Analiza frontului Pareto**

Frontul Pareto obținut este prezentat în Figura 2. În cazul valorii transfigurării consecutive din cadrul regulilor generate, se poate observa cu ușurință că efectul cei mai plauzibil are valoare „medie”. Într-adevăr, regulile cu această etichetă sunt dominante în consecutivă, față de celelalte tipuri de efecte – suport și încredere. Acest fapt se accentuează odată cu creșterea suportului pentru reguli, fără să se manifeste o pierdere semnificativă de încredere în reguli, reprezentativă pentru starea medie de transfigurare. Așadar, se poate deduce că starea cea mai reprezentativă, pentru întreaga bază de date a consumatorilor, este una moderată.

In general terms, flow state is defined as “the process of optimal experience” or the mental state that individuals sometimes experience when they are deeply immersed in certain events, objects or activities (Csikszentmihalyi, 1975, 1977). This concept has been adapted to the Web environment. In this context, flow state is achieved when the consumer is so deeply involved in the process of navigation on the Web that “nothing else seems to matter” (Hoffman & Novak, 1996: p. 57)

Though the model we consider for the experimentation has 12 elements (constructs) interconnected, with 6 fuzzy rule based systems, due to the space constraints, in this paper we focus on that system which considers the four primary antecedents of the consumer’s flow. Specifically, we consider four constructs (speed of interaction, skill/control, challenge/arousal and telepresence/time distortion) as antecedents of the consumer’s flow state (consequent). In this sense, it is been hypothesized that these four elements are positively related to this central construct of the model.

Most parts of the construct, except one of them which was measured by means of an ordinal scale, were gathered by multi-item Likert scales with 9 points; i.e. metric scales. The fuzzy semantic we have applied to all the variables is shown in figure 1.

Training data are composed of 1,154 examples (consumers’ responses). We have run the algorithm 10 times, obtaining the following values for the parameters: 300 generations, size of the population 100, crossover probability 0.7 and the probability of mutation per chromosome 0.1.

#### **4.2. Analysis of the Pareto Front**

The Pareto front we have obtained is shown in Figure 2. With respect to the value taken by the consequent flow in the rules generated, it can be easily observed that the most plausible output is “medium”. Indeed, there is a clear supremacy of the rules with this label in the consequent over the two other outputs in terms of support and confidence. This fact is intensified as the support of the rules grows, without noticing a relevant loss of reliability in the rules which represent medium flow states. Therefore, it can be inferred that the most representative state of flow, for the whole consumers’ database, is moderate.



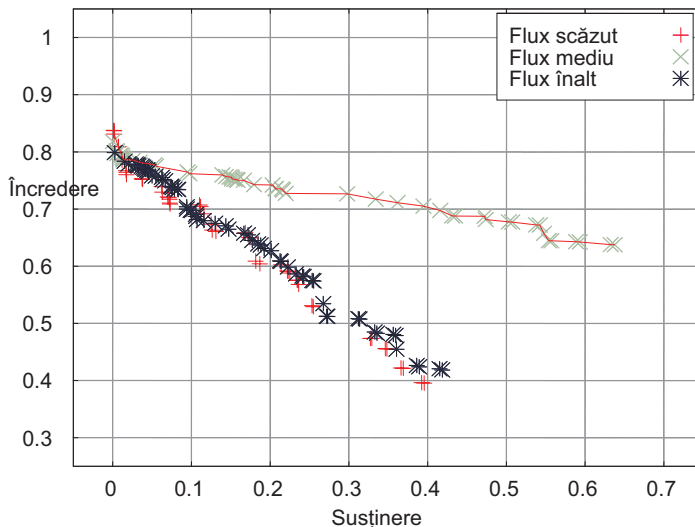


Figura 2. Sub-fronturi Pareto pentru fiecare rezultată a consecutivei, alături de frontul Pareto absolut (cele mai bune reguli din întregul set) legate printr-o linie dreaptă

### 4.3. Analiza ilustrativă a regulilor

Analiza individuală a regulilor generate de această metodă descriptivă este foarte utilă pentru o mai bună înțelegere a comportamentului analizat al consumatorilor. Mai precis, se recomandă o selecție a regulilor din întregul set, pentru alcătuirea frontului Pareto absolut, ținând cont de suportul său (gradul de reprezentativitate a bazei de date a consumatorilor) și, în special, de elementul încredere (gradul de siguranță al tiparului de informație demonstrat de regulă). O selecție ilustrativă a acestor elemente este prezentată în Tabelul 1.

Tabelul 1: Selecție ilustrativă a regulilor deduse din frontul Pareto absolut

	<i>Speed of Interaction</i>		<i>Skill/Control</i>		<i>Challenge/Arousal</i>	<i>Telepresence/Time Distortion</i>	<i>Flow</i>	<i>Sup</i>	<i>Conf</i>
$R_1$	Low	High	Medium			Low	Low	0,0104	0,7980
$R_2$	Medium		Low	High	High	Medium	Medium	0,0102	0,7937
$R_3$	Medium					Medium High	Medium	0,3947	0,7051

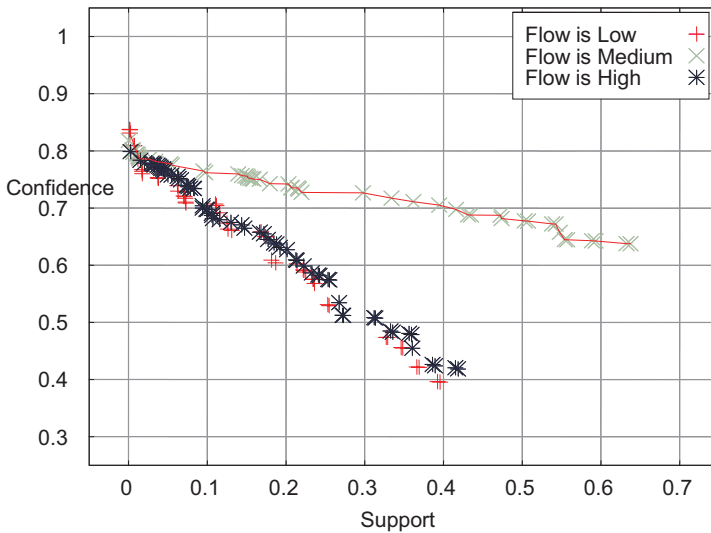


Fig. 2. Sub-Pareto fronts for every output of the consequent, as well as the absolute Pareto front (the best rules from the whole set of rules) joined by a line.

### 4.3. Illustrative Analysis of the Rules

An individual analysis of the rules generated by this descriptive method is very useful to better understand the consumer behavior being analyzed. Specifically, it is recommendable to do a selection of rules from the whole set compounding the absolute Pareto front, paying attention to its support (degree of representativity of the consumers' database) and, especially, to its confidence (degree of reliability of the information pattern shown by the rule). In this regard, we have done an illustrative selection shown in Table 1.

Table 1. Illustrative selection of rules from the absolute Pareto front

	<i>Speed of Interaction</i>		<i>Skill/Control</i>		<i>Challenge/Arousal</i>	<i>Telepresence/Time Distortion</i>	<i>Flow</i>	<i>Sup</i>	<i>Conf</i>
$R_1$	Low	High	Medium			Low	Low	0,0104	0,7980
$R_2$	Medium		Low	High	High	Medium	Medium	0,0102	0,7937
$R_3$	Medium					Medium High	Medium	0,3947	0,7051

În cazul frontului Pareto absolut, R1 este regula cu gradul cel mai înalt de încredere, asociat cu o stare scăzută de transfigurare. În mod similar, R2 reprezintă regula cu gradul cel mai înalt de siguranță dintre cele care reprezintă o stare de transfigurare moderată. În sfârșit, s-a considerat și regula R3, aceea cu cel mai semnificativ suport din întregul set de reguli, la care gradul de încredere este mai mare de 0,7, mai precis, valoarea pragului de încredere stabilit de noi pentru justificarea tiparelor de informații reprezentate prin reguli.

În sinteză, din cele patru antecedente studiate, se reliefează influența percepției referitoare la tele-prezență/distorsiunea de timp (TP/TD) în determinarea stării de transfigurare a consumatorilor; se poate observa modul în care valoarea sa este determinantă în explicarea stărilor de transfigurare redusă (R1) sau moderată (R2 și R3). Într-un mod similar, restul antecedentelor par să exercite o influență slabă sau nulă asupra consecutivei. Acest fapt poate fi determinat și de elementul TP/TD care eclipsează influența tuturor celorlalte. În orice caz, se confirmă afirmația principală extrasă în analiza frontului Pareto, adică inexistența situației în care o combinație de antecedente (reguli) produce stări de transfigurare profunde, cu un nivel semnificativ de siguranță și reprezentativitate. În acest sens, este evident că starea de transfigurare a consumatorilor rămâne moderată în timpul navigării pe internet, chiar și în momentele în care antecedentul cel mai influent (TP/TD) atinge valorile cele mai înalte.

## **5. Observații finale**

Metoda analitică de estimare a modelelor complexe ale comportamentului consumatorilor nu trebuie considerată utilă numai pentru testarea unui set de relații teoretice care compun acest model. Ea trebuie să aibă și capacitatea de a ajuta funcția de management în marketing, în direcția unei perspective exacte asupra anumitor situații de consum, pentru ca ulterior să se ia deciziile cele mai bune. Cercetătorii din domeniul marketingului, mai ales cei care se concentrează asupra îmbunătățirii și dezvoltării „arsenalului” instrumentelor de modelare în marketing, trebuie să fie conștienți de acest fapt, pentru ca propunerile lor să creeze o punte între spațiul academic și cel profesional.

Am prezentat aici o metodologie completă care poate fi aplicată în modelarea cauzală în marketing, cu ajutorul unui sistem genetic aleatoriu, a unui program special de hibridizare, prin metoda inducției descriptive a regulii aleatorii. Această metodă îi permite cercetătorului să-și formeze o imagine complet nouă a relațiilor dintre variabile, în comparație cu metoda folosită pentru determinarea relațiilor din tehnicile statistice specifice domeniului nostru. Se propun tipare unice de informații pentru fiecare relație cauzală din modelul teoretic, ele având rolul de a ghida procesul de învățare de către mașină. Practic, un astfel de proces este susținut de un algoritm genetic și o abordare de optimizare cu obiective multiple, proiectată special pentru un management corespunzător scalelor de măsurare folosite în marketing. În plus, efectele pozitive ale aplicării logicii aleatorii determină exprimarea acestor tipare într-un mod ușor de înțeles pentru mintea umană.

Considering the absolute Pareto front,  $R_1$  is the rule with highest confidence, associated with low states of flow. Likewise,  $R_2$  represents the most reliable rule from those with moderate flow states. Finally, we have also considered the rule  $R_3$ , being the one with highest support among the whole set of rules with confidence higher than 0,7; i.e. the confidence threshold value we have set to give reliability to the information patterns shown by the rules.

Synthetically, from the four antecedents considered, it highlights the influence of the perception about telepresence/time distortion (TP/TD) in determining consumers' states of flow; it can be observed how its value is determinant in explaining low ( $R_1$ ) or moderate ( $R_2$  and  $R_3$ ) states of flow. Likewise, the rest of the antecedents seem to exert a poor or null influence on the consequent. This fact can also be due to the element TP/TD that eclipses the influence of the rest. In any case, it conforms to the main idea we extracted when the Pareto front was analyzed; i.e. a non existence of combinations of antecedents (rules) producing high states of flow, with significant levels of reliability and representativity. In this sense, it is quite illustrative to see how even when the most influential antecedent (TP/TD) takes high values, the consumer's flow state in the process of navigation tends to remain moderate.

## **5. Final Remarks**

An analytical method for estimating complex consumer behavior models should not be only useful to test a set of theoretical relations compounding such model. Moreover, it must be also able to be helpful for the marketing management function to have a good perspective of certain consumption situation, so to take the right decisions. Marketing researchers, especially those focused on improving and developing the "arsenal" of the marketing modeling tools, must be aware of this, in order to bring the gap, with their proposals, between the academics' and the professionals' arenas.

We have presented a complete methodology to be applied in causal marketing modelling by a genetic fuzzy system, a specific soft computing hybridization, with a fuzzy rule descriptive induction approach. This method 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 optimization 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 an easily understandable way regarding the way human beings reason.

## Mulțumiri

Această cercetare a fost parțial sponsorizată de către Ministerul spaniol al Educației și Științei, în cadrul proiectului de cercetare TIN2005-08386-C05-01. O versiune mai scurtă a fost prezentată la Conferința EMAC 2007.

## Bibliografie

- Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Casillas, J., Cordon, O., Herrera, F., Magdalena, L. (Eds) (2003a). *Interpretability issues in fuzzy modeling*. Springer, Heidelberg, Germany.
- Casillas, J., Cordon, O., Herrera, F., Magdalena, L. (Eds) (2003b). *Accuracy improvements in linguistic fuzzy modeling*. Springer, Heidelberg, Germany.
- Casillas, J., Martínez-López, F.J., & Martínez, F.J. (2004). Fuzzy association rules for estimating consumer behaviour models and their application to explaining trust in Internet shopping. *Fuzzy Economic Review*, IX(2), 3-26.
- Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology*, 15(3), 41-63.
- Csikszentmihalyi, M. (1977). *Beyond boredom and anxiety*. Second edition. San Francisco: Jossey-Bass.
- Deb, K., Pratap, A., Agarwal, S., & Meyarevian, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Computation*, 6(2), 182-197.
- Dubois, D., Prade, H., Sudkamp, T. (2005). On the representation, measurement, and discovery of fuzzy associations. *IEEE Trans. Fuzzy Systems*, 13(2), 250-262.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, S., Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. M.I.T. Press.
- 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*, 17, 209-214.
- Hoffman, D., Novak, T. (1996). Marketing in hypermedia computer-mediated environments: conceptual foundations. *Journal of Marketing* 60(July), 50-68.
- Korzaan, M.L. (2003). Going with the flow: predicting online purchase intentions. *Journal of Computer Information Systems*, Summer, 25-31.
- 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.
- Lindskog, P. (1997). Fuzzy identification from a grey box modeling point of view. In Hellendoorn & Driankov (Eds.), *Fuzzy model identification*, pp. 3-50. Springer-Verlag, Heidelberg, Germany.

## Acknowledgments

This research has been supported in part by the Spanish Ministry of Education and Science under research project TIN2005-08386-C05-01. A shorter version of this paper was presented at the 2007 EMAC Conference.

## References

- Ajzen, I., & Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Casillas, J., Cordon, O., Herrera, F., Magdalena, L. (Eds) (2003a). *Interpretability issues in fuzzy modeling*. Springer, Heidelberg, Germany.
- Casillas, J., Cordon, O., Herrera, F., Magdalena, L. (Eds) (2003b). *Accuracy improvements in linguistic fuzzy modeling*. Springer, Heidelberg, Germany.
- Casillas, J., Martínez-López, F.J., & Martínez, F.J. (2004). Fuzzy association rules for estimating consumer behaviour models and their application to explaining trust in Internet shopping. *Fuzzy Economic Review*, IX(2), 3-26.
- Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology*, 15(3), 41-63.
- Csikszentmihalyi, M. (1977). *Beyond boredom and anxiety*. Second edition. San Francisco: Jossey-Bass.
- Deb, K., Pratap, A., Agarwal, S., & Meyarevian, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Computation*, 6(2), 182-197.
- Dubois, D., Prade, H., Sudkamp, T. (2005). On the representation, measurement, and discovery of fuzzy associations. *IEEE Trans. Fuzzy Systems*, 13(2), 250-262.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, S., Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. M.I.T. Press.
- 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*, 17, 209-214.
- Hoffman, D., Novak, T. (1996). Marketing in hypermedia computer-mediated environments: conceptual foundations. *Journal of Marketing* 60(July), 50-68.
- Korzaan, M.L. (2003). Going with the flow: predicting online purchase intentions. *Journal of Computer Information Systems*, Summer, 25-31.
- 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.
- Lindskog, P. (1997). Fuzzy identification from a grey box modeling point of view. In Hellendoorn & Driankov (Eds.), *Fuzzy model identification*, pp. 3-50. Springer-Verlag, Heidelberg, Germany.

- Luna, D., Peracchio, L.A., De Juan, M.D. (2002). Cross-cultural and cognitive aspects of Web site navigation. *Journal of the Academy of Marketing Science*, 30(4), 397-410.
- Novak, T., Hoffman, D., Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology*, 13 (1/2), 3-16.
- Novak, T., Hoffman, D., Yung, Y. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19(1), 22-42.
- Roberts, J.H. (2000). The intersection modelling potential and practice. *International Journal of Research in Marketing* 17, pp. 127-134.
- Steenkamp, J., Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International Journal of Research in Marketing*, 17, 195–202.
- Sugeno, M., Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems* 1(1), pp. 7-31.
- Van Bruggen, G.H., Wierenga, B. (2000), "Broadening the perspective on marketing decision models", *International Journal of Research in Marketing*, vol. 17, pp. 159-168.
- Wind, J. (2006). Challenging the mental models of marketing. In J.N. Sheth and R.S. Sisodia (Eds.) *Does marketing need reform? Fresh perspectives on the future*, M.E. Sharpe, pp. 91-104.
- Witten, I.H., 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., Frank, E. (2000), *Data mining: practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann Publishers, San Francisco, CA, USA.
- Zadeh, L.A. (1965). Fuzzy sets, *Information and Control* 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* 8, 8, 9, pp. 199-249, 301-357, 43-80.

- Luna, D., Peracchio, L.A., De Juan, M.D. (2002). Cross-cultural and cognitive aspects of Web site navigation. *Journal of the Academy of Marketing Science*, 30(4), 397-410.
- Novak, T., Hoffman, D., Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology*, 13 (1/2), 3-16.
- Novak, T., Hoffman, D., Yung, Y. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19(1), 22-42.
- Roberts, J.H. (2000). The intersection modelling potential and practice. *International Journal of Research in Marketing* 17, pp. 127-134.
- Steenkamp, J., Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International Journal of Research in Marketing*, 17, 195–202.
- Sugeno, M., Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems* 1(1), pp. 7-31.
- Van Bruggen, G.H., Wierenga, B. (2000), "Broadening the perspective on marketing decision models", *International Journal of Research in Marketing*, vol. 17, pp. 159-168.
- Wind, J. (2006). Challenging the mental models of marketing. In J.N. Sheth and R.S. Sisodia (Eds.) *Does marketing need reform? Fresh perspectives on the future*, M.E. Sharpe, pp. 91-104.
- Witten, I.H., 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., Frank, E. (2000), *Data mining: practical machine learning tools and techniques with Java implementations*, Morgan Kaufmann Publishers, San Francisco, CA, USA.
- Zadeh, L.A. (1965). Fuzzy sets, *Information and Control* 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* 8, 8, 9, pp. 199-249, 301-357, 43-80.