Consumer Modeling by Multiobjective Genetic Fuzzy Systems: A descriptive rule induction approach

Submitting author: Francisco J. Martínez-López
Affiliation: Associate Professor in Marketing
OtherAuthors: Jorge Casillas
Affiliations: Professor in Computer Science and Artificial Intellingence

Abstract. This paper is the result of an interdisciplinary research between the marketing and the artificial intelligence fields. It 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, knowledge discovery methodology, genetic fuzzy systems, consumer's behavior patterns.

Track: Interdisciplinary Research in Marketing

Consumer Modeling by Multiobjective Genetic Fuzzy Systems: A descriptive rule induction approach

1. Introduction

The field of Knowledge Discovery in Databases (KDD) has lots of potential to support current marketing decision problems. Several academics have recently noted this question, when emphasizing the logical evolution that marketing modeling methods must describe towards systems based on Artificial Intelligence and KDD methodologies (Shim et al., 2002; Wedel, Kamakura & Böckenholt, 2000). Our work in the last years has aimed to contribute to the rapprochement of these fields. Specifically, this paper presents a KDD methodology developed *ad hoc* to be applied in marketing (causal) modeling. A *descriptive rule 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. In this respect, two questions may arise, whose answers are convenient at this introductory section: why fuzzy rules? and, why genetic algorithms (GAs)? In other words, why use these tools of representation and learning instead of others widely used in KDD?

The use of fuzzy rules (instead of interval rules, decision trees, etc.) is mainly justified by the type of data we work with (see section 2.1). In our case, each element/construct of the marketing model is determined/measured by a set of indicators (observed variables) which give partial information to describe it. This adds uncertainty to the data that it can be easily treated with fuzzy rules. Also, it is possible to express the available knowledge of a marketing expert by means of linguistic semantics. Finally, fuzzy rules obtained present high legibility, an important question in KDD.

With respect to the use of GAs to induce fuzzy rules instead of other machine learning techniques, it is due to the following aspects. On the one hand, as the quality of the different fuzzy rules is valued by contradictory objectives –such as support and confidence–, we opt for a multiobjective optimization to treat them adequately. This is currently one of the alternatives with more potential, as well as one of the signs of identity, in AGs, where it stands out due to its superior performance when compared with other techniques. Furthermore, to achieve higher compacity, thus interpretability, we consider a flexible representation of the fuzzy rules which can be easily handled with GAs.

The paper is structured as follows. Section 2 introduces our KDD methodology proposal (a brief extract). In section 3 we empirically apply the methodology on a consumer model. Then, some rules are commented on to illustrate the kind of results we can obtain by this methodology. Finally, we give some concluding remarks.

2. Consumer Behavior Modeling with Fuzzy Rules: A Knowledge Discovery Methodology

2.1. 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 point of the KDD process, so to give suitable and adapted solutions to the specific data we find in consumer behavior modeling. As the *partial interpretation philosophy* is the measurement approach currently predominant in the marketing modeling discipline (Steenkamp & Baumgartner, 2000), we will take it into account when facing how process the data; such approach poses to jointly consider multiple indicators –imperfect when considered individually, though reliable when considered altogether– of the subjacent construct to obtain valid measures

2.2 Data Processing

Next, it is necessary to adapt the collected data to a scheme easily tractable by fuzzy rule learning methods. Though there are several options that can be used to pre-process the data in order to adapt the observed variables to a fuzzy rule learning method (see, for more detail: Casillas, Martínez-López & Martínez, 2004), we propose a more sophisticated process that allows working with the original format without any pre-processing stage: the *multi-item fuzzification.* 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 2.4.2.

2.3. Representation and Inclusion of Expert Knowledge

2.3.1. 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. 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. There are several options, broadly summarized in linear vs. non linear. In this respect, we pose that it is more appropriate to use linear functions, inasmuch as it facilitates the latter interpretation of relations; and
3. The *membership function shapes* should also be fixed. We define the membership function shapes where, given the set $S=\{min,\ldots,max\}$ defining the interval, they hold the following condition:

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

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

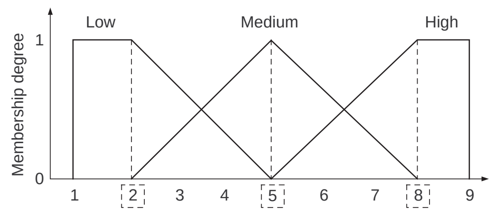

Fig. 1. Fuzzy semantic from a transformation of a 9-point marketing scale (rating scale).

2.3.2. 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.

2.4. 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 Section 1, 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, i.e. support and confidence. To jointly consider both criteria, we propose the use of *multiobjective GAs* The next section describes the main elements of this method we propose.

2.4.1. 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 $X_i,\ i \in \{1, \ldots, n\}$ takes 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

$\min\{1, a+b\}$ 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.

2.4.2. Multi-item Fuzzification

In order to properly consider the set of indicators available for each input/output variable (as discussed in Section 2.2), 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 $\vec{y} = (y_1, \ldots, y_t, \ldots, y_q)$, respectively, the fuzzy propositions X_i is $\tilde{A}_i$ and Y is B are respectively interpreted as follows:

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

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

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

2.4.3. 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 [100 011|2].

2.4.4. 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:

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

where N is the size of the database (the sample size or number of respondents), $\mathsf{x}^{(e)} = (\vec{x}_1^{(e)}, \ldots, \vec{x}_n^{(e)})$ and $\vec{y}^{(e)}$ is the *e*th instance multi-item of input and output respectively, T the *product* T-norm, and $\mu_A(\mathsf{x}^{(e)}) = \min_{i\in\{1,\ldots,n\}} \mu_{\tilde{A}_i}(\vec{x}_i^{(e)})$

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 2.4.2 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:

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

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

2.4.5. 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.

2.4.6. 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).

3. Results of the experimentation and interpretation

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. As the authors allow the use of their database for academic purposes, we have opted for experimenting our methodology with a consumer model already validated and widely known by the academics. This is a plausible and orthodox alternative, as we can see by analyzing other research previously developed (see, as e.g.: Beynon, Curry & Morgan, 2001; Fish et al., 2004; Hurtley, Moutinho & Stephens, 1995; Levy & Yoon, 1995; Rhim & Cooper, 2005)

3.1 Some theoretical notes about the model used for the experimentation

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.

3.2. 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.

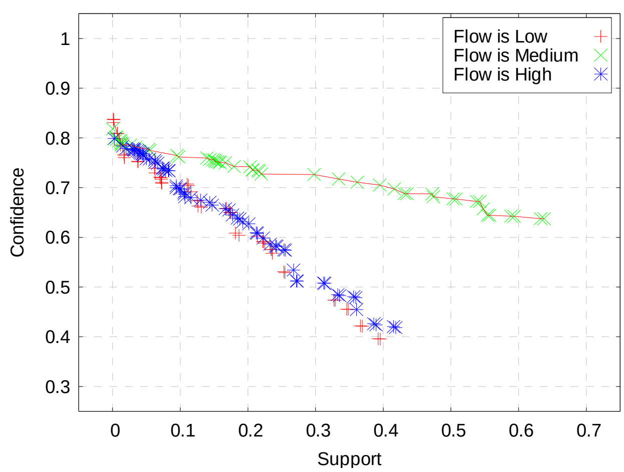

Fig. 2. Sub-Pareto fronts for every output of the consequent; the absolute Pareto front is joined by a line.

3.3. 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

	Speed of Interaction	*Skill/Control*	*Challenge/ Arousal*	*Telepresence/ Time Distortion*	*Flow*	*Sup*	*Conf*
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

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. Concluding remarks

We have faced an interesting problem of KDD in relation to marketing causal modeling and its resolution by genetic fuzzy systems. The problem presents a specific type of data with uncertainty which justifies the use of fuzzy rules. Furthermore, we have practised a multi-objective optimization in order to obtain rules with high degrees of support and confidence. The KDD methodology proposed is being successfully applied to a real problem of consumer behavior in online environments.

In our research agenda, we have the use of other metrics such as consistency and interest of the rules. Also, the unsupervised learning of fuzzy association rules, i.e. without using any antecedent or consequent previously fixed by the marketing expert.

References

Ajzen, I., & Fishbein, M.(1980). Understanding Attitudes and Predicting Social Behavior. Englewood Cliffs, NJ: Prentice-Hall,Inc.

Beynon, M., Curry, B., & Morgan, P. (2001). Knowledge discovery in marketing. An approach through rough set theory. *European Journal of Marketing*, 35(7/8), 915-935.

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.

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.

Fish, K.E., Johnson, J.D., Dorsey, R.E., & Blodgett, J.G. (2004). Using an artificial neural network trained with a genetic algorithm to model brand share. *Journal of Business Research*, 57 (1), 79-85.

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.

Hurley, S., Moutinho, L., & Stephens, N.M. (1995). Solving marketing optimization problems using genetic algorithms. *European Journal of Marketing*, 29(4), 39-56.

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

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

Novak, T., Hoffman, D., Yung, Y. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, 19(1), 22-42.

Rhim, H., & Cooper, L.G. (2005). Assessing potential threats to incumbent brands: New product positioning under price competition in a multisegmented market. *International Journal of Research in Marketing*, 22, 159-182.

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

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

Wedel, M.; Kamakura, W.; Böckenholt, U. (2000). Marketing data, models and decisions. *International Journal of Research in Marketing*, 17, 203-208.