Mathware & Soft Computing 16 (2009) 29-43

Unsupervised Learning of Fuzzy Association Rules for Consumer Behavior Modeling

Albert Orriols-Puig[1], Jorge Casillas[2], and Francisco J. Martínez-López[3]

[1] Grup de Recerca en Sistemes Intel·ligents,
Enginyeria i Arquitectura La Salle (URL), 08022 Barcelona (Spain).
e-mail: aorriols@salle.url.edu

[2] Department of Computer Science and Artificial Intelligence,
Universidad de Granada, 18071 Granada (Spain).
e-mail: casillas@decsai.ugr.es

[3] Dpto. Comercialización e Investigación de Mercados,
Universidad de Granada, 18071 Granada (Spain).
e-mail: fjmlopez@ugr.es

Abstract

Marketing-oriented firms are especially concerned with modeling consumer behavior to improve their information and aid their decision processes on markets. For this purpose, marketing experts use complex models and apply statistical methodologies to infer conclusions from data. Recently, the application of machine learning has been detected as a promising approach to complement these classical techniques of analysis. In this paper, we follow this idea and propose a system, addressed as Fuzzy-CSar, to extract fuzzy association rules from certain consumption problem analyzed. But, as a differentiating sign of identity from other methods, Fuzzy-CSar does not assume any aprioristic causality (so model) within the variables forming the consumer database. Instead, the system is responsible for extracting the strongest associations among variables, and so, the structure of the problem. Fuzzy-CSar is applied to a real-world marketing problem and the results are compared with those obtained by a multi-objective genetic fuzzy system expressly designed for this marketing problem. The results show the advantages of evolving fuzzy association rules and the competitiveness of Fuzzy-CSar in general.

1 Introduction

Companies are constantly searching for suitable marketing opportunities to survive in increasingly turbulent and volatile markets. For this purpose, marketing experts are especially concerned with the creation and management of key information about the market [5]. In management and marketing disciplines, the use of models has been usual to drive the database analysis. Model-based analytical processes imply that a structure of relations among the elements (i.e., variables)

of this previously known model be used to, by means of analytical methods of study, describe or predict the behavior of those relations. This analytical approach matches the procedure classically set by the scientific method; i.e., a researcher works with a set of hypotheses of expected relationships among variables, those hypotheses are empirically tested and, finally, some conclusions are extracted (e.g., see [10]). Basically, these are the core questions in marketing modeling, which are usually followed to drive the information search process in marketing databases with the aim of supporting marketing decisions. But, would it be plausible to work without models? Doubtless, models are very necessary, though their usage may limit the added-value extracted from the data when applied to certain kind of decision problems in marketing. In particular, in non- or ill-structured problems, analysis based on the *a priori* information offered by a model, which may disregard important relationships due to the weak structure of the problem, may not be as effective as a decision maker would expect. In these situations, the so-called knowledge-based marketing support systems could be of great utility (see [2]) . In this regard, several authors have proposed to apply supervised machine learning methods, which are informed with little prior knowledge about the problem, resulting in the extraction of key knowledge that was not detected by the classical analysis methodology (e.g., see [3, 9]). Continuing with these efforts, the application of unsupervised learning techniques which have no knowledge about the problem structure—letting the machine extract interesting, useful, and unknown knowledge about the market—appears as an appealing approach to these problems.

The purpose of this paper is to propose the extraction of fuzzy association rules to discover new interesting knowledge from marketing databases. Specifically, we focus on a database that contains information about the consumer behavior. To achieve this, we design Fuzzy-CSar, a learning classifier system (LCS) [7] that assumes no structure about the problem and evolves a diverse set of fuzzy association rules that describe interesting associations among problem variables. Fuzzy-CSar uses a fuzzy representation that enables the system to deal with the imprecision of the marketing data. The system is compared with the *genetic cooperation-competition learning* GCCL approach that extracts fuzzy rules that define a particular prefixed output variable [9]. The results highlight that fuzzy association rules permit extracting key knowledge that was discovered neither by the classical approach nor by the GCCL approach.

The paper is organized as follows. Section 2 describes the type of data found in marketing problems, explains the classical marketing analysis approach in more detail, and motivates the use of machine learning to tackle these problems. Section 3 provides the basic concepts of association rules, and Sect. 4 describes Fuzzy-CSar. Section 5 presents the experimental methodology and Sect. 6 analyzes the results. Finally, Sect. 7 concludes and provides further work.

2 Consumer Behavior Modeling

A common practice in consumer behavior modeling, when working with complex models, is specifying such models to be empirically analyzed by structural equation

modeling [9]. These models are compounded by elements (constructs) which are inferred from imprecise data, i.e., the indicators or variables related to every element of the model. As follows, we explicate these types of problems, specifically focusing on the type of data that is made available for analysis. Then, we outline some significant aspects related to this structural modeling methodology when applied to a consumer behavior model and motivate the use of machine learning techniques to obtain new interesting information. Finally, we discuss different strategies to let machine learning techniques deal with the particularities of the marketing data.

2.1 Data Collection in Marketing

Generally, when working with complex models for consumer behavior analysis, so with structural models, the elements of the model are divided into two categories: (1) *unobserved/latent* variables, also know as *constructs*, which are conceptually those whose measurement can not be made directly with a single measure; and (2) observed variables or indicators, those related to every single measure (i.e., an item in a multi-item measurement scale) developed to be related to a construct. The underlying idea is that an observed variable is an imperfect measure of a construct, but a set of indicators related to a construct, considered altogether, may lead to a reliable measurement of said construct. Therefore, every construct in a model is usually related to a set of observed variables. This is currently the predominant measurement approach, known as the partial-interpretation philosophy [11].

Finally, there is an especial category of constructs known as second-order constructs. These are characterized by not having direct association with indicators in the measurement model, as an ordinary/first order construct has, but by being defined by a combination of first-order construct related to them. Note that the overall structure of these data is unconventional. Thus, machine learning techniques need to be adapted to deal with them.

2.2 The Classical Approach to Deal with Marketing Data

To extract key knowledge from the data collected by questionnaires, marketing experts use the following approach, addressed as the *classical approach* of analysis in the rest of this paper. First, the expert establishes a structural model of the data, which denotes the relationships—and directions of these relationships—among the variables of the problem. Marketing experts create these models from *a priori* information of the market and from their own experience. Then, the models are used to establish a set of hypotheses that explain the relationship among constructs that have been connected in the structural model. Thereafter, statistical methods based on structural modeling methodologies are used to contrast these hypotheses. The conclusions extracted from the analysis may cause the marketing expert to refine the structural model and to apply again the same analysis procedure.

While it has been shown that the classical approach may provide key knowledge of the consumer behavior analyzed, which may be used to support decision making [10], to be based on a conceptual/structual model to drive the search of information in the database may hamper the discovery of some key knowledge. To extract

further interesting information, several authors have successfully applied machine learning techniques to these types of problems. For example, in [3], the authors used supervised learning techniques to model the consumer behavior in the Internet, resulting in new interesting knowledge not detected by the classical approach. This approach permitted extracting fuzzy rules that also predicted the same variable in the consequent. In the present paper, we take some of the ideas presented in [3] as starting point and extend them to build a system that extracts fuzzy association rules from consumer behavior databases. Thence, we do not consider any *a priori* information about the system and expect that the system provides us with any relevant association among variables. Before proceeding with the description of this approach, the next subsection briefly discusses how a general learning system can be adapted to deal with the particularities of the marketing data.

2.3 Application of Machine Learning to the Marketing Data

In general, two strategies could be used to let learners deal with the marketing data: (1) preprocessing the input data to render them tractable with a general learner or (2) adapting the learning technique to the particularities of the data. The former approach implies transforming the data to a simpler format. An intuitive approach would be to reduce the different items of a specific first-order construct to a single value (e.g., by averaging the values); a similar approach should be used to get an average value for second-order constructs. Another approach would be to expand any variable measured by multiple items to multiple variables measured by a single item and do not consider the existence of second-order constructs; then, the data set could be reduced by means of instance selection.

Nevertheless, the underlying problem of data preprocessing is that relevant information may be lost in the transformation process. For this purpose, Casillas and Martínez-López [3] proposed a modification of the inference process of fuzzy rule-based systems to deal with this especial type of data, which was addressed as *multi-item fuzzification*. The idea of this approach is to use fuzzy operators to (1) aggregate by fuzzy unions (T-conorms) the information provided by the multiple items that define a single variable and (2) intersect (with T-norms) the partial information provided by the first-order variables that describe second-order variables. This mechanism, included in our system, is detailed in Sect. 4.2.

3 Mining Fuzzy Association Rules

Mining association rules consists in extracting rules that denote interesting associations among variables. These rules take the form $X \longrightarrow Y$, where X and Y are sets of variables [1]. The interestingness of the rules is usually evaluated with two measures: (1) the support (*supp*), an estimate of the number of examples in which the association appears with strength and (2) the confidence (*conf*), an estimate of the strength of the association. For example, in a supermarket basket data, association rules such as "if ham then cheese" with supp=20% and conf=90% can be extracted, indicating that a customer that buys ham usually buys cheese also

(conf=90%) and that this association occurs in 20% of the cases.

Several approaches has been proposed to extract association rules from data described by continuous attributes. Among them, the incorporation of fuzzy logics into association rules has been one of the most appealing approaches (e.g., see [6, 8]). In this case, variables are defined by fuzzy sets allowing the system to extract rules such as "if *experience* is large then *income* is high," where large and high are two fuzzy sets. This approach is further explained in the following section, where Fuzzy-CSar is presented in detail.

4 Description of Fuzzy-CSar

Fuzzy-CSar is a machine learning technique that combines *genetic algorithms* (GAs) and apportionment of credit algorithms to evolve a population of fuzzy association rules online. In what follows, we first present the knowledge representation and the mechanisms employed to deal with the particularities of the marketing data, that is, the multi-item fuzzification. Finally, we explain the learning organization.

4.1 Knowledge Representation

Fuzzy-CSar evolves a *population* [P] of *classifiers*, where each classifier consists of a *fuzzy association rule* and a set of parameters. The fuzzy association rule is represented as

$$\textbf{if } x_i \text{ is } \widetilde{A_i^k} \text{ and } \cdots \text{ and } x_j \text{ is } \widetilde{A_j^k} \textbf{ then } x_c \text{ is } \widetilde{A_c^k},$$

in which the antecedent contains a set of ℓ_a input variables $x_i, \ldots, x_j$ ($0 < \ell_a < \ell$, where ℓ is the number of variables of the problem) and the consequent consists of a single variable x_c which is not present in the antecedent. Note that we allow rules to have an arbitrary number of variables in the antecedent. Each variable is represented by a disjunction of *linguistic terms* or *labels* $\widetilde{A_i^k} = \{ A_{i1} \vee \ldots \vee A_{in_i} \}$. To avoid creating largely general rules, which may provide poor information to human experts, the system permits the configuration of the maximum number of linguistic terms per input variable (maxLabIn) and output variable (maxLabOut).

Each classifier has six main parameters: (1) the support *supp*, an indicator of the occurring frequency of the rule; (2) the confidence $conf$, which denotes the strength of the implication; (3) the fitness F, which is computed as a power of the confidence, so reflecting the quality of the rule; (4) the experience *exp*, which counts the number of times that the antecedent of the rule has matched an input instance; (5) the numerosity *num*, which reckons the number of copies of the classifier in the population; and (6) the average size of the association sets *as* in which the classifier has participated.

4.2 Multi-item Fuzzification

In [9], the authors proposed the concept of multi-item fuzzification to deal with data in which each variable was described by a set of items. This procedure, which

was incorporated into Fuzzy-CSar, considers both (1) how to compute the matching degree of a set of items with a variable and (2) how to calculate the matching of several first-order variables with a second-order variable.

The first idea of the method is that each individual item provides partial information about the corresponding first-order variable. Therefore, the matching degree is computed as the union (T-conorm) of the information given by each item. Thence, the matching degree of a variable i with the vector of items $\vec{x_i} = (x_1^i, x_2^i, \ldots, x_{p_i}^i)$ is $\mu_{\widetilde{A_i}}(\vec{x_i}) = max_{h_i=1}^{p_i} \mu_{\widetilde{A_i}}(x_{h_i}^i)$, where we consider the maximum as the union operator.

In addition, the multi-item fuzzification considers the variables of second order are represented by the intersection of the variables of first order. Thence, it computes the matching degree as the T-norm of the matching degrees of each first-order variable. In our implementation, we used the minimum as T-norm.

4.3 Process Organization

Fuzzy-CSar incrementally learns from a stream of examples. That is, at each learning iteration, Fuzzy-CSar receives an input example (e_1, e_2, ..., e_ℓ) and takes the following actions to incrementally update the classifier's parameters and to discover new promising rules. First, the system creates the *match set* [M] with all the classifiers in the population that match the input example with a degree larger than 0. If [M] contains less that θ_{mna} classifiers, the *covering operator* is triggered to create as many new matching classifiers as required to have θ_{mna} classifiers in [M]. Then, classifiers in [M] are organized in *association set candidates*.

Each association set candidate is given a probability to be selected that is proportional to the average confidence of the classifiers that belong to this association set. The selected *association set* [A] goes through a *subsumption* process which aims at diminishing the number of rules that express similar associations among variables. Then, the parameters of all the classifiers in [M] are updated. At the end of the iteration, a GA is applied to [A] if the average time since its last application is greater than θ_{GA}. This process is repeatedly applied, therefore, updating the parameters of existing classifiers and creating new promising rules online.

To completely understand how the system works, five elements need further explanation: (1) the covering operator, (2) the procedure to create association set candidates, (3) the association set subsumption mechanism, (4) the parameter update procedure, and (5) the rule discovery by means of a GA. In the following, each of these elements is explicated in more detail.

Covering Operator. Given the sampled input example e, the covering operator creates a new classifier that matches e with maximum degree. That is, for each variable, the operator randomly decides (with probability $1 - P_\#$) whether the variable has to be in the antecedent of the rule, with the constraints (1) that, at least, a variable has to be selected and (2) that, at most, $\ell - 1$ variables can be included in the antecedent. Then, one of the remaining variables is selected to be in the rule consequent. Each of these variables is initialized with the linguistic label that maximizes the matching degree with the corresponding input value.

In addition, we introduce generalization by permitting the addition of any other linguistic term with probability $P_{\#}$, with the restrictions that each variable in the antecedent and consequent respectively contains *maxLabIn* and *maxLabOut* linguistic terms at maximum.

Creation of Association Set Candidates. The aim of creating association set candidates or niches is to group rules that express similar associations to establish a competition among them and so let the best ones take over their niche. For this purpose, Fuzzy-CSar uses the following approach, which relies on the idea that rules that have the same variable with the same or similar linguistic terms in the consequent must belong to the same niche, since probably they would denote similar associations among variables. First, Fuzzy-CSar sorts the rules of [M] ascendantly depending on the variable of the consequent. Given two rules r_1 and r_2 that have the same variable in the consequent, we consider that r_1 is smaller than r_2 if $\ell_1 < \ell_2$ or ($\ell_1 = \ell_2$ and $u_1 > u_2$), where ℓ_1, u_1, ℓ_2, and u_2 are the position of first and the last linguistic term of the output variable of each rule respectively.

Once [M] has been sorted, the association set candidates are built as follows. At the beginning, an association set candidate [A] is created and the first classifier in [M] is added to this association set candidate. Then, the following classifier k is added if it has the same variable in the consequent, and ℓ_k is smaller than the minimum u_i among all the classifiers in the current [A]. This process is repeated until finding the first classifier that does not satisfy this condition. In this case, a new association set candidate is created, and the same process is applied to add new classifiers to this association set.

Association Set Subsumption. We designed a subsumption mechanism with the aim of reducing the number of different rules that express the same knowledge. The process works as follows. Each rule in [A] is checked for subsumption with each other rule in [A]. A rule r_i is a candidate subsumer of r_j if it satisfies the following four conditions: (1) r_i has higher confidence and it is experienced enough (that is, $conf^i > conf_0$ and $exp^i > \theta_{exp}$, where $conf_0$ and θ_{exp} are user-set parameters); (2) all the variables in the antecedent of r_i are also present in the antecedent of r_j (r_j can have more variables in the antecedent than r_i); (3) both rules have the same variable in the consequent; and (4) r_i is more general than r_j. A rule r_i is more general than r_j if all the input and the output variables of r_i are also defined in r_j, and r_i has, at least, the same linguistic terms than r_j for each one of its variables.

Parameter Update. At the end of each learning iteration, the parameters of all the classifiers that belong to the match set are updated. First, the experience of the classifier is incremented. Second, the support of each rule is updated as

$$supp_{t+1} = \frac{supp_t \cdot (\ell time - 1) + \mu_{\widetilde{A}}(x^{(e)}) \cdot \mu_{\widetilde{B}}(y^{(e)})}{\ell time}, \tag{1}$$

where $\ell time$ is the life time of the classifier, that is, the number of iterations that the classifier has been in the population, and $\mu_{\widetilde{A}}(x^{(e)})$ and $\mu_{\widetilde{B}}(y^{(e)})$ are the matching degree of the antecedent and the consequent with $x^{(e)}$ and $y^{(e)}$ respectively. Then,

the confidence is computed as $conf_{t+1} = sum\_imp_{t+1}/sum\_mat_{t+1}$, where

$$sum\_imp_{t+1} = sum\_imp_t + \mu_{\tilde{A}}(x^{(e)}) \cdot max\{1 - \mu_{\tilde{A}}(x^{(e)}), \mu_{\tilde{B}}(x^{(e)}), \text{ and} \quad (2)$$

$$sum\_mat_{t+1} = sum\_mat_t + \mu_{\tilde{A}}(x^{(e)}). \quad (3)$$

Initially, $sum\_imp_{t+1} = sum\_mat_{t+1} = 0$. Next, the fitness of each rule in [M] is computed as $F = conf^{\nu}$, where ν permits controlling the pressure toward highly fit classifiers. Finally, the association set size estimate of all rules that belong to the [A] is updated. Each rule maintains the average size of all the association sets in which it has participated.

Discovery Component. The GA is triggered on [A] when the average time from its last application upon the classifiers in [A] exceeds the threshold θ_{GA}. It selects two parents p_1 and p_2 from [A], where each classifier has a probability of being selected proportional to its fitness. The two parents are crossed with probability P_χ, generating two offspring ch_1 and ch_2. Fuzzy-CSar uses a uniform crossover operator that contemplates the restriction that any offspring has to have, at least, a variable in the rule's antecedent. If crossover is not applied, the children are a exact copy of the parents. The resulting offspring may go through three different types of mutation: (1) mutation of antecedent variables (with probability $P_{I/R}$), which randomly chooses whether a new antecedent variable has to be added to or one of the antecedent variables has to be removed from the rule; (2) mutation of the linguistic terms of the variable (with probability P_μ), which selects one of the existing variables of the rule and mutates its value; and (3) mutation of the consequent variable (with probability P_C), which selects one of the variables of the antecedent and exchanges it with the variable of the consequent. Thereafter, the new offspring are introduced into the population. If the population is full, excess classifiers are deleted from [P] with probability directly proportional to their association set size estimate and inversely proportional its fitness.

5 Problem Description and Methodology

Having described Fuzzy-CSar, now we are in position to start with the experimentation. In what follows, we first explicate the characteristics of the data and provide a previous model that was extracted from these data by means of classical marketing analysis. Then, we present the experimental methodology.

5.1 Problem Description

The problem addressed in this paper is the modeling of web consumers to extract key knowledge that enable marketing experts to create a compelling online environment for these users. To tackle this problem, several authors have proposed causal models of the consumer experience on the Internet [4]. These models have mainly focused on the description of the state of *flow* during consumer navigation of the Web, that is, the cognitive state experienced during online navigation. Reaching the state of flow comprises a "complete involvement of the user with his activity,"

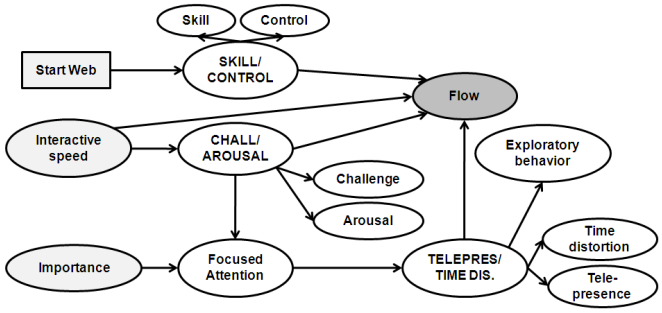

Figure 1: Theoretical model of the user experience on the Internet.

and so, marketing experts are especially concerned with identifying the factors that lead the user to the state of maximum flow. As follows, we present one of the most influential of these structural models [10], which is taken as starting point in this paper; besides, we also explain how the data used in our experiments was collected.

Figure 5.1 depicts the structural model analytically developed by Novak et al. [10], which was described by nine *constructs*: skill, control, interactive speed, importance, challenge, arousal, telepresence, time distortion, and exploratory behavior. Some of these first-order constructs were used to partially define *second-order constructs*—see, for example, that skill and control partially define the *skill/control* variable. In addition, the model also considered the variable *startWeb*, which indicated for how much time the user had been using the web.

The data were obtained by means of the surveys used in [10]. These surveys proposed a set of questions or *items* that partially described each one of the nine first-order constructs. The user was asked to grade these questions with Likert nine-point rating scales that ranged from "strongly disagree" to "strongly agree." The startWeb variable was measured with a six-point rating scale that comprised different options of usage time.

The analysis conducted in [10] identified that the following four constructs were the most important ones to determine the state of *flow*: (1) skill and control, (2) challenge and arousal, (3) telepresence and time distortion, and (4) interactive speed. The other constructs were found to be meaningless to define *flow*. However, it is worth noting that the conclusions extracted by the classical approach depended on the initial causal model. Therefore, some key relationships may had not been discovered. For this reason, the use of other techniques that do not employ any *a-priori* knowledge about the problem, such as Fuzzy-CSar, appears as an appealing approach to discover new valuable knowledge about the consumer.

5.2 Experimental Methodology

In order to perform a careful analysis of the capabilities and the added value provided by Fuzzy-CSar, the system was applied to three different subsets of the marketing data. These three experiments are explicated as follows.

Experiment 1. In the first experiment, we forced that all rules had the *flow* variable as output and considered the four input variables that were found to be relevant in Novak's et al. model: (1) skill and control, (2) challenge and arousal, (3) telepresence and time distortion, and (4) interactive speed. The purpose of this experiment was to study whether Fuzzy-CSar could extract association rules that were equivalent to the associations denoted in Novak's et al. causal model.

Experiment 2. The second experiment was analogous to the first one, with the exception that all the input variables were considered. The aim of this experiment was to examine whether, with the inclusion of the variables that were considered useless by Novak's et al. model, we could obtain new interesting knowledge.

Experiment 3. In the last experiment, we analyzed whether we could obtain further relevant knowledge by letting the system evolve rules with any variable in the consequent. Thence, in this case, we were searching for any interesting association among variables instead of focusing only on the *flow* variable.

In all the experiments, Fuzzy-CSar was configured with a population size of 6 400 rules and the following parameters: $P_{\#} = 0.5$, $P_{\chi} = 0.8$, $\{P_{I/B}, P_{\mu}, P_{C}\} = 0.1$, $\theta_{GA} = 50$, $\theta_{exp} = 1\,000$, $conf_0 = 0.95$, $\nu = 1$, $\delta = 0.1$. All the variables, except for *startWeb*, used Ruspini's strong fuzzy partitions with three linguistic terms. *startWeb* used six membership functions, each centered in one of the values that the variable could take. In all cases, $maxLabIn = 2$ and $maxLabOut = 1$.

The results of Fuzzy-CSar were compared with those obtained with an evolutionary multi-objective approach based on a GCCL scheme proposed in [3]. This system was expressly designed to evolve a Pareto of fuzzy rules that described a concrete output variable with maximum confidence and support—in the experiments reported in [3], the method was applied to extract rules that described the *flow* construct. The same configuration used by the authors was employed in our experiments; that is, the system was configured to evolve a population of 100 individuals during 100 iterations, with crossover and mutation probabilities of 0.7 and 0.1 respectively. The variables used the same semantics as Fuzzy-CSar ones. The GCCL approach was ran only on the two first experiments. We could not apply the system to the third experiment since it required to fix a unique output variable.

Before proceeding with the experiments, it is worth highlighting the underlying differences between Fuzzy-CSar and the GCCL approach. Fuzzy-CSar learns a set of association rules online. It incrementally updates the parameters of rules and uses an implicit niching mechanism to group similar rules and to apply a GA that aims at maximizing the confidence of the rules. The system indirectly pressures toward obtaining rules with large support since the GA is most often applied to the rules that belong to an association set more frequently. On the other hand, the GCCL approach explicitly optimizes the rules with respect to their support and confidence, that is, it optimizes the Pareto front. Therefore, the GCCL approach is more likely to evolve rules that maximize support and confidence, since it is specifically designed with this objective, while Fuzzy-CSar is more focused on evolving a diverse set of rules that have maximum confidence. Notwithstanding, we are interested in analyzing how our approach performs in comparison with a system which is specialized in optimizing the Pareto front.

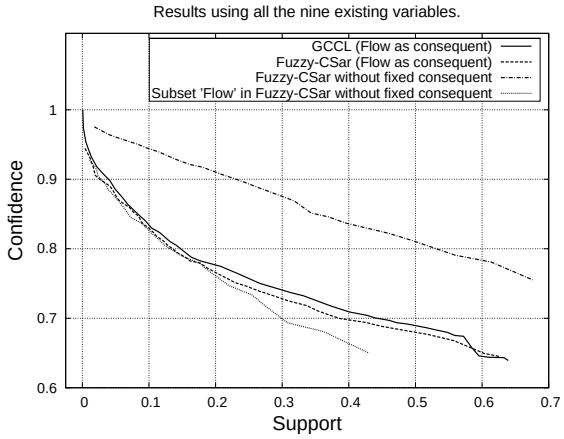

Figure 2: Average Pareto front obtained by Fuzzy-CSar and the GCCL approach considering the 9 variables of the marketing data.

6 Analysis of the Results

This section analyzes the results of the experiments detailed above. First, we study the results from the objective point of view; that is, we only consider the rules that lay in the Pareto set and compare Fuzzy-CSar with the GCCL system. Second, we extend this analysis by examining the particular rules evolved by Fuzzy-CSar and providing some examples that evidence the added value of the association rules.

6.1 Analysis of the Quality of the Rules

As follows, we analyze (1) the experiments that fix the variable *flow* in the consequent of the rules and (2) the experiment where any of the variables can be either in the antecedent or in the consequent.

Experiments Forcing *Flow* in the Consequent. Figure 2 shows the Pareto fronts[1] obtained by Fuzzy-CSar and the GCCL approach on the second experiment, which considers any of the nine variables in the antecedent and forces *flow* to be in the consequent. These Pareto fronts are very similar to those obtained in the first experiment, which considers only the four most relevant variables in the antecedent (the curves are not shown for the sake of brevity). This similarity between the Pareto fronts obtained with both experiments confirms the hypothesis that the four variables identified as the most important ones by the models in [10] are indeed the most relevant ones to describe the *flow* construct.

Table 1 complements the analysis by reporting (1) the total number of rules evolved by Fuzzy-CSar, (2) the number of these rules that are in the Pareto set, and (3) the number of rules in the Pareto sets created by the GCCL approach. For each case, the mean crowding distance between consecutive solutions of the Pareto front

[1]The results correspond to averages of ten runs with different seeds.

Table 1: Average number of rules evolved by Fuzzy-CSar, average number of these rules that are in the Pareto set, and average number of rules in the Pareto sets obtained by the GCCL approach. For each clase, the average crowding distance of the population is provided in parentheses.

	FCSar	**FCSar Par.**	**GCCL Par.**
Experiment 1	479.2 ($2.36 \cdot 10^{-3}$)	76.3 ($1.53 \cdot 10^{-2}$)	82.6 ($1.49 \cdot 10^{-2}$)
Experiment 2	1259.7 ($9.54 \cdot 10^{-4}$)	105.9 ($1.07 \cdot 10^{-2}$)	84.4 ($1.49 \cdot 10^{-2}$)
Experiment 3	1752.5 ($6.88 \cdot 10^{-4}$)	468.3 ($2.58 \cdot 10^{-3}$)	—

are provided in parentheses. The results indicate that the Pareto set evolved by Fuzzy-CSar is very similar to the one created by the GCCL approach. Notice that, in both cases, the solutions are uniformly distributed along the objective space; therefore, both systems are able to find a large variety of non-dominated solutions. The similarity of the results highlight the robustness of Fuzzy-CSar, which is able to generate Pareto fronts that very close to those created by a competent technique which specifically optimizes the Pareto front.

In addition, Fuzzy-CSar also creates a set of rules (479 and 1260 for the first and second experiment on average) which can also be interesting for human experts. That is, Fuzzy-CSar evolves a set of distributed niches which contain rules that are semantically different from the rules that belong to other niches. It is possible that some of the niches contain rules that denote interesting knowledge, but that these rules are dominated by rules of other niches. Then, although these semantically-different dominated rules do not belong to the Pareto set, they may contain key knowledge for marketing experts.

Experiments Permitting any Variable in the Consequent. In the third experiment, we enabled Fuzzy-CSar to evolve any variable in the rules' consequent. Therefore, we let Fuzzy-CSar search for any kind of association among variables with large confidence and support, disregarding the causal models provided by marketing experts. In what follows, we analyze the difference between the Pareto fronts created in this experiment with respect to those obtained in previous experiments.

Figure 2 includes the results of the last experiment, enabling the comparison with the Pareto fronts evolved when the system is forced to create rules with *flow* in the consequent. These results show the potential of our approach. In a single run, Fuzzy-CSar was able to evolve a set of rules with large confidence and support, resulting in a Pareto front that was clearly better than those of Fuzzy-CSar and the GCCL approach when the *flow* construct was fixed in rules' consequent.

To complement these results, the same figure plots the Pareto front evolved by Fuzzy-CSar in this last experiment, but considering only the rules that predict the *flow* construct. Notice that, for large confidence, this Pareto front is close to the one evolved by the GCCL approach and Fuzzy-CSar in previous experiments where *flow* was fixed in the consequent. On the other hand, the Pareto front degrades as the confidence of the rules decreases. This behavior can be easily explained as follows. As the number of possible variables in the consequent increases, Fuzzy-CSar needs to maintain a larger number of rules that belong to different niches.

In this case, the implicit niching system together with the niche-based GA and population-wise deletion operator of Fuzzy-CSar make pressure toward maintaining a diverse set of solutions. On the other hand, the GA also puts pressure toward rules with maximum confidence. Therefore, the system maintains a diverse set of solutions with maximum confidence, which goes in detriment of solutions with smaller confidence, but larger support.

We acknowledge that similar results could be obtained by the GCCL approach by running nine different experiments, each one fixing a different variable in the consequent. This would result in nine different Pareto sets that should be joined and processed to get the final Pareto set. Nevertheless, it is worth noting that Fuzzy-CSar provides a natural support for the extraction of interesting association rules with different variables in the consequent, evolving a set of distributed solutions in parallel, and maintaining only those with maximum confidence.

6.2 Examples of the Added Value of Fuzzy-CSar Rules

After showing the competitiveness of Fuzzy-CSar with respect to the GCCL approach, this section analyzes the importance of the knowledge provided by some of the rules discovered by Fuzzy-CSar. For this purpose, we show two particular examples of rules that provide key knowledge considered neither by the structural model [10] nor by the GCCL approach [3].

First, we selected a rule that predicted *exploratory behavior*, that is,

R_1: **IF** *importance* **is** Medium **and** *skill/control* **is** {Small or Medium} **and** *focusedAttention* **is** {Small or Medium} **and** *flow* **is** {Small or Medium} **THEN** *exploratoryBehavior* **is** Medium [Supp.: 0.22; Conf.: 0.87].

The model proposed by Novak et al. considered that *exploratory behavior* was related to only *telepresence/time distortion*, that is, the degree of telepresence and the effect of losing the notion of time while browsing the web. However, rule R_1 does not consider this relationship. Instead, it denotes that *exploratory behavior* is determined by *importance*, perceived *skill/control*, *focused attention* in the browsing process, and the state of *flow*. Thence, this rule indicates that intermediate values of the variables of the antecedent explicate, with confidence 0.87, states of moderate exploratory behaviors in the Web. The knowledge denoted by the rule may cause the marketing expert to consider other associations among variables that were not considered in the initial model. In particular, this relationship was initially considered in the causal model built in [10], but it was further discarded after a process of model refinement. Nonetheless, R_1 is alerting of the importance and strength of this association.

Second, we chose the following rule, which described *focused attention*:

R_2: **IF** *importance* **is** {Small or Medium} **and** *chall/arousal* **is** {Small or Medium} **and** *telepres/time distortion* **is** Medium **and** *exploratoryBehavior* **is** {Medium or Large} **THEN** *focused attention* **is** Medium [Supp.: 0.21; Conf.: 0.84]

In Novak's et al. model, *focused attention* was related to neither *importance* nor *chall/arousal*. However, rule R_2 indicates that these two variables together with

telepres/time distortion and *exploratory behavior* may determine moderate degrees of attention in the Web browsing. This information is especially interesting since it contradicts the causal model. This contradiction is reasonable if we consider the following. Differently from [10], Fuzzy-CSar does not assume any type of problem structure. Thence, Fuzzy-CSar can discover new relations among variables that may appear to be very useful and interesting. This may be the case of R_2, which implies that increasing the experience in the navigation process may influence, together with the other variables, the capacity of users to focus their attention on the Web. In summary, R_2 proposes a new scenario that was not considered before, and marketing experts may analyze whether this new knowledge needs to be included in further revisions of the causal model.

In addition to these particular examples, it is worth emphasizing that, in general, unsupervised learning techniques such as Fuzzy-CSar may be relevant tools in problems for which *a priori* information is unknown. In these cases, association rules may discover interesting, useful, and hidden associations among variables that help marketing experts build a causal model.

7 Conclusions and Further Work

In this paper, we discussed the importance of applying machine learning techniques to marketing problems—specifically, to the modeling of the consumer behavior—in order to complement the knowledge extracted by the classical analysis approaches. Differently from the classical approach, which requires that marketing experts provide a structural model, we proposed not to use any type of *a priori* information and to use a system to discover this structural model in form of association rules. For this purpose, we employed Fuzzy-CSar, a general-purpose unsupervised learning technique that evolves a set of association rules online and that uses adapted inference mechanisms to deal with the particularities of the marketing data.

The empirical results showed the robustness of Fuzzy-CSar. Specifically, Fuzzy-CSar was able to evolve non-dominated solutions with support and confidence values similar to those created by a multi-objective GCCL approach when the output variable was fixed. In addition, when not imposing any constraint to the output variable, Fuzzy-CSar evolved Pareto sets that were by far superior to those obtained by the GCCL approach.

Along with the discussion of the results, we already hypothesized that some of the rules that were not included in the Pareto sets may also provide interesting knowledge. In line with this observation, in further work, it would be interesting to analyze the diversity in the populations evolved by Fuzzy-CSar in more detail. The reported results indicate that the system is able to evolve a large variety of rules, and only a few proportion of these rules form the Pareto front. Thence, the population has a large number of rules that, despite not being in the Pareto set, can potentially provide human experts with new interesting knowledge. Therefore, further work would focus on the analysis of the semantics of the rules and on the design of methodologies to compare the rules obtained by different systems according to their semantics.

Acknowledgments

The authors would like to thank *Ministerio de Educación y Ciencia* for its support under projects TIN2005-08386-C05-01 and TIN2005-08386-C05-04, *Generalitat de Catalunya* for its support under grants 2005FI-00252 and 2005SGR-00302, and *Junta de Andalucía* for its support under project P07-TIC-3185.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD Int. Conference on Management of Data*, pages 207–216, Washington D.C., May 1993.

[2] G. H. V. Bruggen and B. Wierenga. Broadening the perspective on marketing decision models. *Int. Journal of Research in Marketing*, 17:159–168, 2000.

[3] J. Casillas and F. Martínez-López. Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. *Expert Systems With Applications*, 36(2):1645-1659, 2009.

[4] M. Csikszentmihalyi. Finding flow: The psychology of engagement with everyday life. 1997.

[5] A. Drejer. Back to basics an beyond: Strategic management – an area where practice and theory are poorly related. *Management Decision*, 42(3/4):508–520.

[6] D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery*, 13:167–192, 2006.

[7] J. H. Holland. *Adaptation in natural and artificial systems*. MIT Press, Cambridge, MA., 2nd edition, 1992.

[8] T. P. Hong, C. S. Kuo., and S. C. Chi. Trade-off between computation time and number of rules for fuzzy mining from quantitative data. *Int. Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 9(5):587–604, 2001.

[9] F. Martínez-López and J. Casillas. Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic fuzzy systems. *Industrial Marketing Management*, doi=10.1016/j.indmarman.2008.02.003, 2008.

[10] T. Novak, D. Hoffman, and Y. Yung. Measuring the customer experience in online environments: A structural modelling approach. *Marketing Science*, 19(1):22–42, 2000.

[11] J. Steenkamp and H. Baumgartner. On the use of structural equation models for marketing modelling. *Int. Journal of Research in Marketing*, 17:195–202, 2000.