Automatic Discovery of Potential Causal Structures in Marketing Databases Based on Fuzzy Association Rules

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

[1] Grup de Recerca en Sistemes Intelligents
   Enginyeria i Arquitectura La Salle (URL), 08022 Barcelona, Spain
   e-mail: aorriols@salle.url.edu
[2] CITIC-UGR (Research Center on Communication and Information Technology)
   Department of Computer Science and Artificial Intelligence
   University of Granada, 18071 Granada, Spain
   e-mail: casillas@decsai.ugr.es
[3] Department of Marketing
   University of Granada, 18071 Granada, Spain and
   Universitat Oberta de Catalunya, 08035 Barcelona, Spain
   e-mail: fjmlopez@ugr.es

Abstract. Marketing-oriented firms are especially concerned with modeling consumer behavior in order 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. In the recent years, the application of machine learning has been identified as a promising approach to complement these classical techniques of analysis. In this chapter, we review some of the first approaches that undertake this idea. More specifically, we review the application of Fuzzy-CSar, a machine learning technique that evolves fuzzy association rules online, to a certain consumption problem analyzed. 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 the real-world marketing problem of modeling web consumers, with the aim of identifying interesting relationships among the variables of the model. In addition, the system is compared with a supervised learning technique, which is able to extract associations between a set of input variables and a pre-fixed output variable, expressly designed for this marketing problem. The results show that Fuzzy-CSar can provide interesting information for marketing experts that was not detected by the classical approach, and that the extraction of fuzzy association rules is an appealing alternative, in general, to refine or complement the modeling results obtained with the use of traditional methods of analysis applied for these purposes; in particular, we focus on, and take as a reference, the structural equation modeling.

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 [6]. 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 [20]). 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, especially in the academic field, where the arsenal of statistical and, in general, analytical tools are usually applied with a theory-driven approach. However, mostly from the practitioners' perspective, 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.

Hence, though the support of models is helpful to address the search of information in marketing databases, there are situations, both in the practitioners' and scholars' arena, where the use of other non model-based solutions, either on their own or as a complementary tool to a information search process based on models, might produce profitable results. For instance, from an academic perspective, when analyzing the validity of a theoretical model, an additional approach to the traditional would be to adjust all the possible causal structures (models), reasonable or unreasonable, and then, theoretically analyze those configurations with better fitness. However, as causal (theoretic) structures of reference increase in complexity, the number of possible configurations is considerably higher [3], so the development of the said approach would be more difficult to accomplish. In this case, powerful analytical methods are necessary to undertake this task with efficiency. Time ago, some authors [7] pointed out that a process of search and analysis for all the possible configurations of causal models, in a certain marketing database, could be automatized using some Computation Science-based method. However, these authors also recognized that years of evolution would be necessary to be able to work with suitable procedures.

Nowadays, the so-called knowledge-based marketing support systems offer an excellent framework to develop methods with this purpose (see [3]). 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 [4, 17]). 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 chapter is to review the work done on 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 apply Fuzzy-CSar, a learning classifier system (LCS) [11] 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 an evolutionary multi-objective (EMO) approach that extracts fuzzy rules that define a particular prefixed output variable [17]. The results highlight that fuzzy association rules permit extracting key knowledge that was discovered neither by the classical approach nor by the EMO approach.

The chapter is organized as follows. Section 2 describes the type of data usually found in marketing databases, with especial attention to the particularities of the kind of variables (i.e. constructs) forming complex causal models in marketing, 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 presents future work lines.

2 Previous Considerations on the Adaptation of Marketing Data

A common practice in marketing modeling, and consumer behavior modeling in particular (field where the method proposed here is applied to), when working with complex models (i.e., with multiple relations of dependent and independent variables), is specifying such models to be empirically analyzed by structural equation modeling [17]; other types of causal models, so statistical estimation methods, are also used, though we focus our research on the most difficult case to solve of the complex models. 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. Then, we explain how marketing data can be transformed into fuzzy semantics, and 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 known as *constructs*, which are conceptually those whose measurement cannot 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 [22].

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 constructs 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 a database (usually generated after a survey that administered questionnaires to a sample of the target population), marketing experts use the following approach, addressed as the *classical approach* of analysis in the rest of this chapter. First, the expert establishes a theoretical model, which denotes the relationships—and directions of these relationships—among the variables of the problem. Marketing experts base such models on diverse sources, where we highlight the theoretical basis, the *a priori* information of the market, and 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, a measurement model is set and 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 [20], based on a conceptual/structural model to drive the search of information in the database, it 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 [4], 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 always predicted the same variable

in the consequent. In the present chapter, we take some of the ideas presented in [4] as starting point and extend them to build a system that extracts fuzzy association rules from consumer behavior databases, but with a different approach. In particular, 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 subsections briefly present key questions related with the transformation of the original data (i.e. marketing measurement scales) into fuzzy semantic and finally discuss how a general learning system can be adapted to deal with the particularities of the marketing data.

2.3   Transformation of Marketing Scales into Fuzzy Semantic

The machine learning stage could not work without transforming the original marketing data into fuzzy terms. Some notes are deserved to be commented in this regard. The transformation process differs depending on the type of marketing scale, subjacent to every variable of the marketing database. In order to simplify the problem, let us focus on the following traditional classification of measurement scales [23, 24]: nominal, ordinal, interval, and ratio. The transformation of these basic measurement scales into fuzzy variables is useful for all those cases where a measurement scale entails, as minimum, certain order. This premise would involve all the types of measurement scales, with the exception of the nominal. Next, we offer some general reflections for each of the four scales.

Nominal scales. The characteristics of this scale (e.g., consumer's gender, nationality, etc.) just allow identifying and classifying into some of the categories of the scale. But, there is no relation of order or grade between the categories. Consequently, it does not have any sense applying the fuzzy reasoning, as nature of the scale's categories is purely deterministic. This fact involves that these scales are considered as singleton fuzzy sets, a particular case of fuzzy sets; i.e., if certain consumer belongs to certain category, he/she has a membership degree of one to the fuzzy set related to that category, and zero to the others.

Ordinal scales. When the transformation of these types of scales is tackled, there is a main inconvenient: as they are non-metric/qualitative scales, there is just information about the consumer's membership or non-membership to one of the categories in which the marketing variable was structured. This fact limits the possibilities to determine the extreme and central values of the fuzzy sets defining the linguistic variable associated with that marketing variable.

Likewise, regardless the previous question, the marketing expert should solve the following dilemma: should or should not the linguistic variable explicitly consider the structure of categories defining the original marketing variable? In general, it is widely accepted the convenience of a linguistic variable synthesizes the information provided by the original scale, in order to improve the interpretation of relations among the elements of the model, as well as to draw on the potentials of fuzzy inference. However, a subsequent aggregation of original categories is difficult to

implement due to the lack of information provided by an ordinal scale; i.e., references are needed, for instance the extremes and central points of the fuzzy sets obtained after the aggregation. On the other hand, there are studies which require the original categories of certain ordinal scale to be maintained, with the aim of analyzing, for instance, an eventual research question. Though, it is also true that there are other situations where the aggregation of categories, classes of the variable, can be done without any inconvenient for the research purposes.

Therefore, based on the above reflections, there are two possibilities when transforming an ordinal scale into a linguistic variable: (1) maintaining the categories of the original scale or (2) aggregating such categories, so to obtain a linguistic variable with fewer terms than categories had the original ordinal scale. Diverse questions, mainly the problems that would have to be faced with the latter, make the first option to be more convenient. In Fig. 1 we show an example for the variable "Weekly use of the Web" (ordinal scale extracted from the database used in [20]), structured as follows: (1) $x \leq 1$ hour; (2) $1 < x \leq 5$ hours; (3) $5 < x \leq 10$ hours; (4) $10 < x \leq 20$ hours; (5) $20 < x \leq 40$ hours; and (6) $x > 40$ hours.
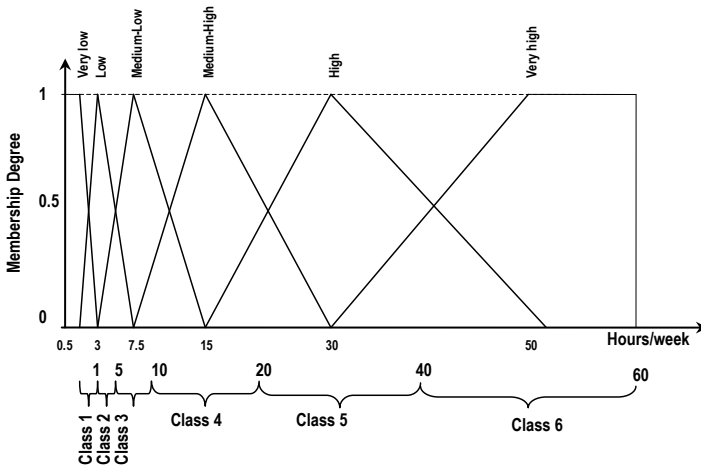

Fig. 1 Example of transformation of a marketing ordinal scale into a linguistic variable (classes of the original scale are maintained)

Interval scales. These scales are metric/quantitative, so they allow more possibilities when being transformed into linguistic variables. Notwithstanding, now we are going to focus our reflections on the particular case of the rating scales, as they are the scales habitually used to measure the items related to constructs. Two main questions should be tackled: the number of linguistic terms to use and the type of membership function more convenient to represent the behavior of the fuzzy sets. With respect to the former, though the particularities of each research has to be taken into account, considering the number of points commonly used by these types of scales (i.e. between five and eleven), it is convenient to work with a number

of fuzzy set between three and five. Likewise, as these scales generally measure the consumer's opinion intensity on variables of interest for the certain research, we propose using, in general, the following labels or terms: *low*, *medium/indifferent*, and *high* when working with three fuzzy sets; and *very low*, *low*, *medium/indifferent*, *high*, and *very high* when working with five fuzzy sets. With respect to the second question, the membership function type, it is more convenient the transformation of scales using a triangular function. In particular, triangular functions must be necessarily used for the case of the extreme fuzzy sets defining a fuzzy variable related to certain rating scale. The main argument to support this is based on the characteristics of these scales.

For instance, let us consider a seven-point semantic differential scale (1: Bad - 7: Good), used to measure the consumer's attitude toward a brand. We know that when consumer has a totally negative attitude, his/her valuation will be 1. However, if his/her valuation were 2, that would mean a low attitude, though not the lowest level linked to a valuation of 1. Therefore, the fuzzy set *low* should show a membership degree of 1 when the marketing scale value is 1, decreasing with a linear tendency to zero for the rest of numeric values associated with said set. This reasoning would be equally valid for the case of the fuzzy set *high*, though it would be a fuzzy set with a membership function linearly increasing till the highest value of 7 in the marketing scale. Finally, as it is logic, the superior limit of the fuzzy set *low*, as well as the inferior limit of the fuzzy set *high*, would match with that value of the marketing scale in which the fuzzy set *medium* takes a membership degree of 1. Therefore, it is also necessary to define the domain of such central fuzzy set of the variable. The procedure to define this set differs depending on whether we are dealing with forced or non-forced scales. For the case of non-forced scales, the solution is intuitive and straight. As the central point of the scale represents an intermediate or indifferent position, said point would be the central point of the fuzzy set, with membership degree of 1. However, the solution for the case of forced-scales would imply another solution, as there is no central point. This fact makes necessary the use of a trapezoidal functions to define the behavior of the central fuzzy set of the variable, in such a way that the central points would be always formed by an even number of points in the scale, i.e., the central points. Figure 2 illustrates a graphic representation of two fuzzy variables with three linguistic terms each, associated with two semantic differential (non-forced and forced) scales.

Ratio scales. These present less restrictions and inconvenient to be transformed into fuzzy variables than any of the other types already described. As these scales are truly continuous, with zero as the lowest value, the numbers of linguistic terms, the determination of the domains for the established set of fuzzy sets, and the membership functions to use are completely flexible. The only inconvenient for the marketing expert is how to fix the maximum value for the scales in order to define the domain of the last fuzzy set of the variable.
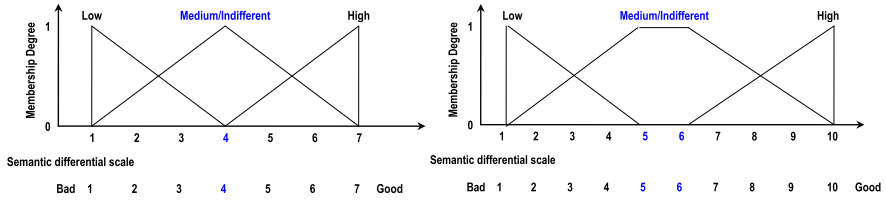
Fig. 2 Examples of membership functions for non-forced (seven-point) and forced (ten-points) rating scales

2.4 *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 into 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 [4] 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 Fuzzy-CSar, is detailed in Sect. 4.2.

3 Mining Association Rules

Association rule mining (ARM) [1] consists in extracting interesting patterns, associations, correlations, or causal structures among the variables in sets of usually unlabeled data. There has been and increasing interest in ARM in the recent years due to the existence of real-world applications in industry that generate large volumes of unlabeled data that have to be processed in order to extract novel and useful information for the company, which in turn may help guide the decision process of the business. This section briefly introduces ARM by first reviewing the initial approaches applied to data described by boolean variables and by then going to more recent approaches that can deal with numeric variables.

3.1 Association Rules: The Beginning

Initial research on ARM was mainly motivated by the analysis of market basket data, which enabled companies to get a better understanding of the purchasing behavior of their customers. Therefore, association rules were first applied to problems featured by binary or boolean variables. The problem of ARM can be described as follows.

Let $T = \{t_1, t_2, ..., t_n\}$ be a set of transactions, where each transaction consists of a set of items $I = \{i_1, i_2, ..., i_k\}$. Let an *itemset* X be a collection of items $I = \{i_1, i_2, ..., i_m\}$. A frequent itemset is an itemset whose support ($supp(X)$) is greater than a threshold specified by the user (this threshold is typically addressed as *minsupp* in the literature). The support of the rules is computed as

$$supp(X) = \frac{|X(T)|}{|T|}. \tag{1}$$

That is, the support is the number of transactions in the database which have the itemset X divided by the total number of transactions in the database.

Then, an *association rule R* is an implication of the form $X \rightarrow Y$, where both X and Y are itemsets and $X \cap Y = \emptyset$. As previously mentioned, ARM aims at extracting interesting association rules. Although many different measures have been developed to measure the interest of association rules so far [15], there are two basic indicators of the quality of the rules: support (*supp*) and confidence (*conf*). The support of a rule is defined as the ratio of the support of the union of antecedent and consequent to the number of transactions in the database, i.e.,

$$supp(R) = \frac{supp(X \cup Y)}{|T|}. \tag{2}$$

The confidence is computed as the ratio of the support of the union of antecedent and consequent to the support of the antecedent, i.e.,

$$conf(R) = \frac{supp(X \cup Y)}{supp(X)}. \tag{3}$$

Therefore, support indicates the frequency of occurring patterns, and confidence evaluates the strength of the implication denoted in the association rule.

Many algorithms have been proposed to extract association rules since the first proposal in [1]. Agrawal et al. [2] presented the Apriori algorithm, one of the most influential algorithms that set the basis for further research in association rule mining. Apriori uses two different phases to extract all the possible association rules with minimum support and confidence: (1) identification of all frequent itemsets and (2) generation of association rules from these large itemsets. The first phase is based on an iterative process that builds k-length itemsets by combining all the (k-1)-length itemsets whose support is greater than or equal to the minimum support fixed by the used. The support of each new itemset is computed by scanning the database. The second phase takes each frequent itemset and generates rules that

contain some of the items in the antecedent of the rule and the remaining ones in the consequent of the rule. The confidence of each rule is computed by scanning the database for each rule, and only those with a minimum confidence are returned. As this process is time consuming, especially in large databases, new approaches that try to reduce the number of scans of the database have been proposed (e.g., see [9]).

3.2 Association Rule Mining with Continuous Variables

The first approaches to ARM only focused on analyzing whether an item was present in a transaction or not, describing the problem with boolean variables. Nonetheless, real-world problems are typically featured by continuous attributes, and these attributes can contain many distinct values. While the support of particular values for these attributes tends to be low, the support of interval of values is much higher. This created the need for building algorithms that could deal with intervals of values, yielding two approaches to the problem: quantitative association rules and fuzzy association rules.

Several authors proposed algorithms to deal with interval-based rules, which are typically addressed as quantitative association rules. In these algorithms, the aim was shifted to extracting rules in which variables are defined by intervals, such as "if *experience* ∈ [5-10] years then *income* ∈ [30 000 - 40 000]$." One of the first methods that falls under this category can be found in [21], which, previously to extracting frequent itemsets, uses an equi-depth partitioning to define a set of intervals for each continuous attribute. The method creates a new variable for each interval, transforming therefore the problem into a binary problem. Then, an Apriori-like algorithm is applied to extract association rules from the transformed data. Although this approach and similar ones could deal with continuous variables, it was detected that these types of algorithms could either ignore or over-emphasize the items that lay near the boundary of intervals if the attributes were not properly partitioned. This was addressed as the *sharp boundary* problem. Two main approaches have been followed to tackle this problem. On the one hand, some authors have applied different clustering mechanisms to extract the best possible intervals from the data [16, 19, 25]. On the other hand, there have been some proposals that adjust these intervals during learning [18].

In parallel to these approaches, some authors faced the problem of sharp boundaries by incorporating fuzzy logics into ARM. In this case, variables were 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 linguistic terms represented by fuzzy sets. As variables were represented by fuzzy sets, the problem of the sharp boundary was overcome. Among others, one of the most significant approaches under this category was proposed in [12, 13, 14], which redefined support and confidence for fuzzy association rules and designed an algorithm that combined ideas of Apriori with concepts of fuzzy sets to extract association rules described by variables represented by linguistic terms.

Since the consumer behavior modeling problem addressed in this chapter is featured by continuous attributes, we employ a system that falls under this last category. Therefore, Fuzzy-CSar is a system that creates fuzzy association rules from the database and utilizes the typical definitions of support and confidence defined for fuzzy systems to evaluate the interestingness of rules. The details of the algorithm are further explained in the following section.

4 Description of Fuzzy-CSar

Fuzzy-CSar is an ARM algorithm that follows a Michigan-style learning classifier system architecture [11] to extract fuzzy association rules from databases. Differently from most of the state-of-the-art algorithms in fuzzy ARM, Fuzzy-CSar (1) uses a fixed-size population to search for the most promising associations among variables, and so, does not necessarily create all the association rules with minimum support and confidence, (2) extracts association rules from streams of examples instead of from static databases, and, as a consequence, (3) does not scan repetitively the data base but incrementally learns from the stream of examples. The system uses an apportionment of credit technique to incrementally adjust the parameters of association rules and a genetic algorithm [8, 10] to discover new promising rules online. In addition, the system is provided with the multi-item fuzzification in order to deal with the particularities of the marketing data. In what follows, the system is described in some detail by first presenting the knowledge representation and the multi-item fuzzification and then explaining the learning organization.

4.1 Knowledge Representation

Fuzzy-CSar evolves a *population* [P] of *classifiers*, where each classifier individually denotes an association among problem variables. Therefore, the solution to the problem is the whole population. Note thus that the population size fixes an upper bound on the number of interesting associations that can be found; that is, at maximum, the system will be able to discover as many interesting relationships as number of classifiers in the population.

Each classifier consists of a *fuzzy association rule* and a set of parameters. The fuzzy association rule is represented as

$$\textbf{if } x_i \text{ is } \widetilde{A}_i \text{ and } \cdots \text{ and } x_j \text{ is } \widetilde{A}_j \textbf{ then } x_c \text{ is } \widetilde{A}_c,$$

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. Thus, we allow rules to have an arbitrary number of variables in the antecedent, but we require that rules have always one variable in the consequent.

Each variable is represented by a disjunction of *linguistic terms* or *labels* $\widetilde{A}_i = \{A_{i1} \vee \ldots \vee A_{in_i}\}$. However, the number of linguistic terms per variable is limited in

order to avoid the creation of largely general rules that may provide poor information about the problem. That is, if no restriction were required, the system would tend to generate rules whose variables in the antecedent and consequent had all the possible linguistic terms, since they would cause the rule to match any possible input, and so, its support and confidence would be very high. To prevent the system from creating these rules, we allow the configuration of the maximum number of linguistic terms permitted per input variable (maxLabIn) and output variable (maxLabOut).

In addition to the rule itself, each classifier has also 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. The function of the different parameters, as well as the process followed to create and evolve these rules, is further explained with the process organization of Fuzzy-CSar in Section 4.3. But before that, next section introduces the multi-item fuzzification included in the system to deal with the marketing data.

4.2   Multi-item Fuzzification

In [17], the authors proposed the concept of multi-item fuzzification in order to deal with problems featured by unobserved variables described by multiple items and second-order constructs partially defined by first-order constructs. This procedure, which was incorporated into Fuzzy-CSar to deal with this kind of marketing data, 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 unobserved variable or first-order variable. Therefore, the authors proposed to compute the matching degree as the aggregation (T-conorm) of the information given by each item. Thence, the matching degree of a variable i with the vector of items $\mathbf{x_i} = (x_1^i, x_2^i, \ldots, x_{p_i}^i)$ is

$$\mu_{\widetilde{A}_i}(\mathbf{x_i}) = max_{h_i=1}^{p_i} \mu_{\widetilde{A}_i}(x_{h_i}^i). \tag{4}$$

In our experiments, we considered the maximum as the union operator.

On the other hand, second-order variables are those defined by the intersection of the information provided by the corresponding first-order variables. For this reason, multi-item fuzzification calculates the matching degree of second-order variables as the T-norm of the matching degrees of each corresponding first-order variable. In our implementation, we used the minimum as T-norm.

4.3 Process Organization

After explaining the classifier representation and the mechanism to compute the matching degree in the marketing data, now we are in position to review the learning organization of Fuzzy-CSar. Fuzzy-CSar incrementally tunes the parameters of the classifiers as new examples are received and periodically applies the GA to *niches* of classifiers in order to create new rules that denote promising associations. The process is explained as follows. At each learning iteration, Fuzzy-CSar receives an input example (e_1, e_2, ..., e_ℓ) and takes the following actions. 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 fully comprehend the system process, 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. The subsequent subsections explicate each one of these elements in more detail.

4.3.1 Covering Operator

The covering operator is the responsible for providing the population with the initial classifiers which will be latter evaluated as new examples are received and evolved by the genetic algorithm. In order to create coherent rules, the operator generates rules that denote associations that are actually strong in the sampled example e from which covering has been activated. For this purpose, the covering operator uses the following procedure. Given the sampled input example e, covering creates a new classifier that contains some variables of the problem in the antecedent and the consequent of the rule and 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.

4.3.2 Creation of Association Set Candidates

The system organizes the population rules in different niches that individually contain rules with similar associations with the aim of establishing a collaboration among niches and a competition of rules inside each niche. That is, the collaboration/competition scheme is produced by the niche-based genetic algorithm and the population-based deletion scheme, which are explained in subsection 4.3.5. The following explains the heuristic process employed to create these niches.

The system 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. Therefore, in order to create the different association set candidates, Fuzzy-CSar first 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, the system considers 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. At the end, this process creates a non-fixed number of niches and distributes the rules through these niches.

4.3.3 Association Set Subsumption

The system explained thus far may generate similar rules that would coexist in the population. In order to avoid the maintenance of similar rules in the population, which would consume resources that may be useful to discover rules that denote different associations, Fuzzy-CSar incorporates a subsumption mechanism that searches for similar rules and only maintains the most general one.

The subsumption procedure 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 as r_j for each one of its variables.

4.3.4 Parameter Update

At the end of each learning iteration, the parameters of all the classifiers that belong to the match set are adjusted according to the information provided by the sampled instance. First, the experience of the classifier is incremented. Second, the support of each rule is updated as

$$supp_{t+1} = \frac{supp_t \cdot (\ell time - 1) + \mu_{\widetilde{A}}(x^{(e)}) \cdot \mu_{\widetilde{B}}(y^{(e)})}{\ell time}, \tag{5}$$

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 degrees of the antecedent and the consequent with $x^{(e)}$ and $y^{(e)}$ respectively. Note that this formula computes the support considering all the examples sampled to the system since the rule was created.

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

$$sum\_imp_{t+1} = sum\_imp_t + \mu_{\widetilde{A}}(x^{(e)}) \cdot max\{1 - \mu_{\widetilde{A}}(x^{(e)}), \mu_{\widetilde{B}}(y^{(e)})\}, \text{ and} \tag{6}$$

$$sum\_mat_{t+1} = sum\_mat_t + \mu_{\widetilde{A}}(x^{(e)}). \tag{7}$$

Initially, $sum\_imp_{t+1} = sum\_mat_{t+1} = 0$. That is, sum_imp maintains the addition of the matching degree of each example sampled so far with the implication of the rule, and sum_mat keeps the addition of the matching degrees of the antecedent of the rule with each example sampled since the rule creation.

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

4.3.5 Discovery Component

Fuzzy-CSar uses a niche-based GA to create new promising classifiers. The GA is triggered on [A] when the average time from its last application upon the classifiers in [A] exceeds the threshold θ_{GA}. The time elapsed between GA applications enables the system to adjust the parameters of the new classifiers before the next application of the GA.

Once triggered, the GA 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 exact copies 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.

To sum up, Fuzzy-CSar is a population-based ARM that evaluates rules online as new examples are sampled to the system and that periodically applies a GA to create new promising rules. Note that the system does not require the user to determine the minimum support and minimum confidence of the rules. Instead of this, the system evolves a set of rules with maximum support and confidence, and the number of rules is limited by the population size. The rule activation based on matching prioritizes rules that match a larger number of training examples with respect to those that match a lower number of training examples. In addition, the confidence-based selection of [A] and the inside-niche competition established by the GA pressure toward the creation of rules with progressively higher confidence.

5 Problem Description and Methodology

After motivating the use of ARM for modeling the user behavior and presenting a technique that is able to extract fuzzy association rules without specifying the minimum support and the minimum confidence of the rules, we now move on to the experimentation. This section first explains the details of the marketing problem analyzed in this chapter and presents previous structural models extracted from this problem by using classical marketing analysis techniques. Then, we detail the experimental methodology.

5.1 Problem Description

The present work addresses the problem of modeling web consumers to extract key knowledge that enable marketing experts to create a compelling online environment for these users with the final goal of using this information to create a competitive advantage on the Internet. To tackle this problem, several authors have proposed causal models of the consumer experience on the Internet [5]. 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
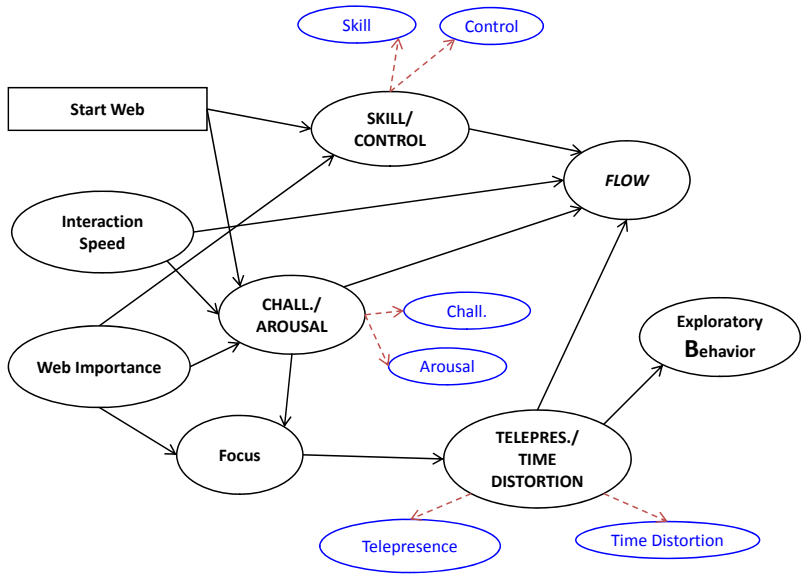
Fig. 3 Theoretical model of the user experience on the Internet

the state of flow comprises a "complete involvement of the user with his activity." Therefore, marketing experts are especially concerned with the identification of the factors that lead the user to the state of maximum flow.

In this chapter, we consider one of the most influential structural models as starting point, which was analytically developed by Novak et al. [20]. The structural model, illustrated in Fig. 3, consisted of nine *first-order constructs*: skill, control, interactive speed, importance, challenge, arousal, telepresence, time distortion, and exploratory behavior. In addition to these first-order variables, the model also contained three *second-order constructs*: skill/control, chall/arousal, and telepresence/time distortion, which are partially defined by first-order constructs. The model also considered the variable *startWeb*, which indicated for how long the user had been using the web.

The data were obtained from a large sample Web-based consumer survey conducted in [20]. These surveys posed 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-ordinal rating scale that comprised different options of usage time.

The analysis conducted in [20] 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. In the following, we use the model-free system Fuzzy-CSar in order to identify associations among the variables of the problem with the aim of detecting any further relationship not captured by the causal model of Novak et al. The approach is not proposed as an alternative to the classical marketing analysis tools, but as a complement to these techniques. The next subsection details the experiments conducted.

5.2 Experimental Methodology

The aim of the experiments was to study whether the application of machine learning techniques could result in the identification of not only the same but new important associations between variables with respect to those detected in the causal model of Novak et al. In addition, we also analyzed the benefits of association rule mining over other machine learning techniques for data prediction, i.e., techniques in which the target variable is predetermined. For this purpose, we included an EMO approach expressly designed for creating rules with a fixed variable in the consequent for the marketing problem [4]. The approach used a genetic cooperative competitive scheme to evolve a Pareto set of rules with maximum support and confidence. For more details on the algorithm, the reader is referred to [4]. Herein, we explain the three experiments conducted to analyze the added value provided by ARM.

Experiment 1. The first experiment aimed at studying whether Fuzzy-CSar could capture the same knowledge represented in the structural model of Novak et al. This model focused on predicting the variable *flow*. The analytical study detected that there were four relevant variables to determine the state of *flow*: (1) skill and control, (2) challenge and arousal, (3) telepresence and time distortion, and (4) interactive speed. The remaining variables were considered irrelevant. Thus, we applied Fuzzy-CSar to the data extracted from the questionnaires, but only considering these four variables and fixing the variable *flow* as the output variable of the association rules. As the output variable was fixed, we could also apply the aforementioned EMO approach in order to analyze whether Fuzzy-CSar could obtain similar rules to those created by a system specialized to optimize the support and confidence of the rules.

Experiment 2. Since the first experiment did not consider all the input variables, some important knowledge could be overlooked by no considering important interactions between these missing variables. To study this aspect, we ran Fuzzy-CSar on the data described by all the input variables and compared the results with those obtained from the first experiment. Again, the variable *flow* was fixed as the target variable of the association rules. In addition, we also run the EMO approach on

these data, extending the comparison of Fuzzy-CSar and the EMO approach started in the previous experiment.

Experiment 3. The two first experiments permitted the analysis of whether the machine learning techniques could extract similar knowledge to that provided by the structural model of Novak et al. and whether new important knowledge was discovered. Nevertheless, the two first experiments did not test the added value provided by extracting association rules online. Therefore, in the third experiment, we ran Fuzzy-CSar on the input data without forcing any variable in the consequent. Thus, the system was expected to evolve rules with different variables in the consequent, and so, to evolve the rules with maximum support and confidence. The aim of the experiment was to examine whether new interesting relationships, not captured by the structural model, could be discovered by Fuzzy-CSar. Note that, since the output variable was not fixed in this experiment, the EMO approach could not be ran.

In all the experiments, Fuzzy-CSar was configured with a population size of $6\,400$ rules and the following parameters: $P_\# = 0.5$, $P_\chi = 0.8$, $\{P_{I/B}, P_\mu, P_C\} = 0.1$, $\theta_{GA} = 50$, $\theta_{exp} = 1\,000$, $conf_0 = 0.95$, $v = 1$, $\delta = 0.1$. 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* $= 1$. For the EMO approach, we employed the same configuration used by the authors [4]. 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.

Before proceeding to the analysis of the results, it is worth highlighting the underlying differences between Fuzzy-CSar and the EMO approach. As aforementioned, the first important difference is in the knowledge representation: Fuzzy-CSar creates fuzzy association rules where the output variable is not fixed and the EMO approach creates rules with a prefixed target variable. Therefore, Fuzzy-CSar, and ARM algorithms in general, could create rules that denote important associations among variables in a single run; on the other hand, the EMO approach has to fix the output variable at each run. The second important difference is the process organization and the goal of the method. Fuzzy-CSar aims at learning a set of association rules distributed through different niches according to the genotype of the rule; in addition, the learning is done online. The fitness-based inside-niche selection and population-based deletion pressure toward obtaining rules with maximum confidence and support. Conversely, the EMO approach explicitly optimizes the rules with respect to their support and confidence, that is, it optimizes the Pareto front. Therefore, the EMO 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.

6   Analysis of the Results

This section examines the results of the three experiments from two perspectives. First, we study the results on the objective space by analyzing the rules of the Pareto set, that is, those rules for which there do not exist any other rule in the population that has both a higher support and a higher confidence than the given rule. With this analysis we consider the ability of Fuzzy-CSar to create different rules with high support and confidence that are distributed through the solution space and compare it with the EMO approach, but we do not study the utility of the rules from the point of view of the marketing expert. This analysis is conducted afterwards, where several rules that provide new interesting knowledge about the problem, not captured by the structural model, are illustrated.

6.1   Analysis of the Rules in the Objective Space

Figure 4 shows the shape of the Pareto fronts evolved by Fuzzy-CSar and by the EMO approach in the first experiment, which considers only the four most relevant variables in the antecedent and forces *flow* to be in the consequent. The first row of Table 1 complements this information by reporting the average number of rules in the population of Fuzzy-CSar and the average number of rules in the Pareto set of Fuzzy-CSar and the EMO approach. In addition, to indicate the distribution of solutions in the Pareto set, the sum of the distance crowding between consecutive solutions in the Pareto front are also provided in parentheses.
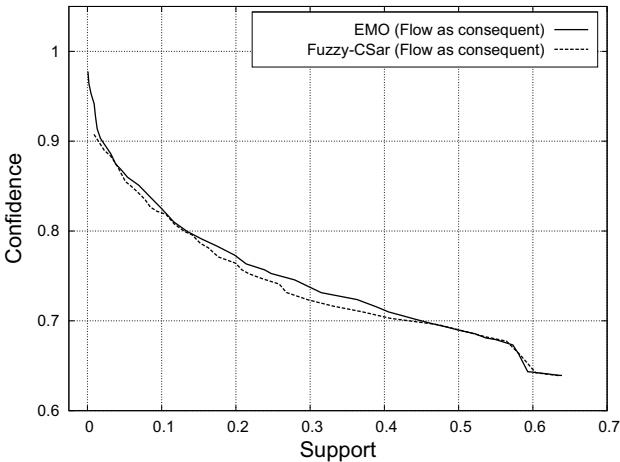

Fig. 4 Average Pareto front obtained by Fuzzy-CSar and the EMO approach considering the 4 variables of the marketing data identified as the most important variables by the structural model and fixing the variable *flow* as target of the rules
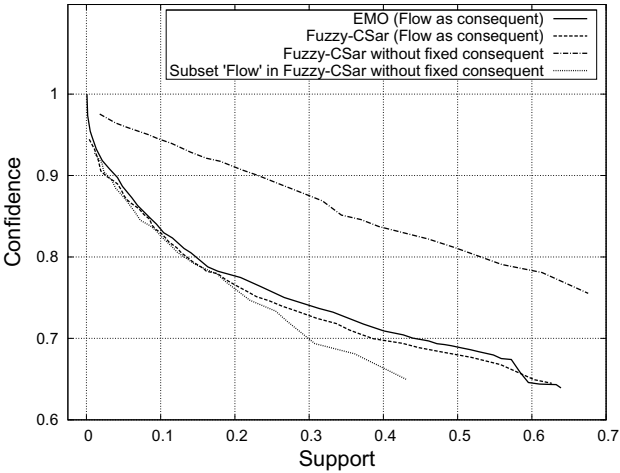
Fig. 5 Average Pareto front obtained by Fuzzy-CSar and the EMO approach considering the 9 variables of the marketing data

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 EMO approach. For the Pareto sets, the average crowding distance of the population is provided in parentheses.

	FCSar All	FCSar Pareto	EMO Pareto
Experiment 1	479.2	76.3 $(1.53 \cdot 10^{-2})$	82.6 $(1.49 \cdot 10^{-2})$
Experiment 2	1259.7	105.9 $(1.07 \cdot 10^{-2})$	84.4 $(1.49 \cdot 10^{-2})$
Experiment 3	1752.5	468.3 $(2.58 \cdot 10^{-3})$	—

These results show the similarity of the results obtained with both methods. That is, both techniques discovered a similar number of solutions in the Pareto set, and these solutions were distributed uniformly through the objective space. The similarity of the results highlight the robustness of Fuzzy-CSar, which was able to generate Pareto fronts that were very close to those created by a competent technique which specifically optimized the Pareto front. In addition, the strongest rules obtained denote the same relationships provided by the structural model. Nonetheless, we put aside further discussion about the utility of the rules until the next section.

Figure 5 together with the second row of Table 1 show the same information but for 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. Actually, the EMO approach can discover practically the same rules than those obtained with the first experiment. On the other hand, Fuzzy-CSar obtains a significantly larger number of rules in the Pareto set; as a consequence, the average crowding distance decreases, since

solutions in the Pareto set are closer to each other. Nonetheless, the shape of the Pareto sets is almost the same in both cases, which supports the hypothesis that the four variables identified as the most important ones by the models in [20] are indeed the most relevant ones to describe the *flow* construct.

Finally, Fig. 5 together with the third row of Table 1 supply the results of Fuzzy-CSar for the third experiment, where any variable can be in the antecedent or in the consequent of the association rules. 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 EMO approach when the *flow* construct was fixed in the rule consequent.

To complement the results of the third experiment, the same figure plots the objective values of the rules of the Pareto front evolved by Fuzzy-CSar that predict the *flow* construct. Notice that, for large confidence, this Pareto front is close to the one evolved by the EMO approach and Fuzzy-CSar in previous experiments where *flow* was fixed in the consequent. On the other hand, the solutions in the Pareto front degrade 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 pressures toward evolving 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.

Similar results could be obtained by the EMO approach by running nine different experiments, each one fixing a different variable in the consequent. This would yield nine Pareto sets, each one with rules that predict one of the nine variables. Then, these Pareto sets could 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 *Analysis of the Utility of the Rules from the Marketing Expert Perspective*

After showing the competitiveness of Fuzzy-CSar with respect to the EMO 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 [20] nor by the EMO approach [4].

Firstly, 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 [20], but it was further discarded after a process of model refinement. Nonetheless, R_1 is alerting of the importance and strength of this association.

Secondly, 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 [20], 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 the variables forming a database that help marketing experts better understand a certain problem they are approaching to.

7 Conclusions and Further Work

This chapter started by discussing the importance of the use of machine learning techniques to give support to classical methodologies for marketing analysis. Among the different techniques in machine learning, we identified ARM as one of the most appealing approaches since it enables the automatic identification of associations or relationships among variables from a data set. That is, differently from the classical approach, which requires that marketing experts work with a theoretical model, ARM does not require any type of *a priori* information about the problem.

In order to show the added value that ARM could provide to marketing experts, we reviewed the application of Fuzzy-CSar, a general-purpose ARM technique that evolves a set of association rules online and that uses adapted inference mechanisms to deal with the particularities of the marketing data. Then, we applied it to the problem of modeling the user navigational process in online (the Web) environments; in particular, we were based on the Novak et al. [20] data and *flow* model to develop the experimental stage. Additionally, the system was compared to a predictive EMO approach that needed to fix the target variable of the evolved rules. The empirical results highlighted the added value of applying machine learning techniques to the marketing problem and, more specifically, of extracting association rules. That is, Fuzzy-CSar was able not only to generate rules that expressed the same knowledge as that contained in the theoretical (structural) marketing model of reference, but also to capture additional relationships among variables not previously considered in the theoretical background. We have shown how some of such uncovered relationships are very interesting from the analyzed marketing problem perspective. To sum up, these results suggest the suitability of ARM for marketing databases analysis. In particular, it has demonstrated to be helpful in consumer behavior modeling, especially as a complementary analytical tool to the traditional methods applied there. Anyhow, marketing researchers and practitioners, especially the formers, must not forget that the outcomes of these new, less orthodox, analytical methods are desirable to be interpreted and assimilated without forgetting to connect with the subjacent theoretical frameworks of the marketing issues they face.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington D.C., May 1993, pp. 207–216 (1993)

2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, September 1994, pp. 487–499 (1994)
3. Bollen, K.A.: Structural equations with latent variables. Wiley-Interscience, Hoboken (1989) (A division of John Wiley & Sons, Inc.)
4. Casillas, J., Martínez-López, F.J.: Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. Expert Systems With Applications 36(2), 1645–1659 (2009)
5. Csikszentmihalyi, M.: Finding flow: The psychology of engagement with everyday life (1997)
6. Drejer, A.: Back to basics an beyond: Strategic management – an area where practice and theory are poorly related. Management Decision, 42(3/4), 508–520
7. Glymour, C., Scheines, R., Spirtes, P., Kelly, K.: Discovering causal structure. Academic Press, Orlando (1987)
8. Goldberg, D.E.: Genetic algorithms in search, optimization & machine learning, 1st edn. Addison Wesley, Reading (1989)
9. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery 8(1), 53–87 (2004)
10. Holland, J.H.: Adaptation in natural and artificial systems. The University of Michigan Press, Ann Arbor (1975)
11. Holland, J.H., Reitman, J.S.: Cognitive systems based on adaptive algorithms. In: Waterman, D.A., Hayes-Roth, F. (eds.) Pattern-directed inference systems, pp. 313–329. Academic Press, San Diego (1978)
12. Hong, T.P., Kuo, C.S., Chi, S.C.: A fuzzy data mining algorithm for quantitative values. In: Proceedings International Conference on Knowledge-Based Intelligent Information Engineering Systems, pp. 480–483 (1999)
13. Hong, T.P., Kuo, C.S., Chi, S.C.: Mining association rules from quantitative data. Intelligent Data Analysis 3, 363–376 (1999)
14. Hong, T.P., Kuo, C.S., Chi, S.C.: Trade-off between computation time and number of rules for fuzzy mining from quantitative data. International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems 9(5), 587–604 (2001)
15. Lenca, P., Meyer, P., Vaillant, B., Lallich, S.: On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid. European Journal of Operational Research 184, 610–626 (2008)
16. Lent, B., Swami, A.N., Widom, J.: Clustering association rules. In: Procedings of the IEEE International Conference on Data Engineering, pp. 220–231 (1997)
17. Martínez-López, F.J., Casillas, J.: Marketing intelligent systems for consumer behaviour modelling by a descriptive induction approach based on genetic fuzzy systems. Industrial Marketing Management (2009), doi:10.1016/j.indmarman.2008.02.003
18. Mata, J., Alvarez, J.L., Riquelme, J.C.: An evolutionary algorithm to discover numeric association rules. In: SAC 2002, pp. 590–594. ACM, New York (2002)
19. Miller, R.J., Yang, Y.: Association rules over interval data. In: SIGMOD 1997: Proceedings of the 1997 ACM SIGMOD international conference on Management of data, pp. 452–461. ACM, New York (1997)
20. Novak, T., Hoffman, D., Yung, Y.: Measuring the customer experience in online environments: A structural modelling approach. Marketing Science 19(1), 22–42 (2000)

21. Srikant, R., Agrawal, R.: Mining quantitative association rules in large relational tables. In: Jagadish, H.V., Mumick, I.S. (eds.) Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Quebec, Canada, pp. 1–12 (1996)
22. Steenkamp, J., Baumgartner, H.: On the use of structural equation models for marketing modelling. International Journal of Research in Marketing 17, 195–202 (2000)
23. Stevens, S.S.: On the theory of scales of measurement. Science, 677–680 (1946)
24. Stevens, S.S.: Measurement, psychophysics and utility. John Wiley, New York (1959)
25. Wang, K., Tay, S.H.W., Liu, B.: Interestingness-based interval merger for numeric association rules. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, KDD, pp. 121–128. AAAI Press, Menlo Park (1998)