Journal of Business Research 66 (2013) 1332–1337

Contents lists available at SciVerse ScienceDirect

Journal of Business Research

A soft-computing-based method for the automatic discovery of fuzzy rules in databases: Uses for academic research and management support in marketing

Albert Orriols-Puig [a,*], Francisco J. Martínez-López [b,e], Jorge Casillas [c], Nick Lee [d]

[a] Research Group in Intelligent Systems, Ramon Llull University, Spain
[b] Dept. of Management, University of Granada, Spain; and Marketing Group, Open University of Catalonia, Spain
[c] Dept. of Computer Science and Artificial Intelligence & CITIC-UGR, University of Granada, Spain
[d] Marketing Department, Aston University, UK
[e] Zicklin School of Business, The City University of New York, USA

ARTICLE INFO

Article history:
Received 1 April 2011
Received in revised form 1 September 2011
Accepted 1 November 2011
Available online 12 July 2012

Keywords:
KDD
Unsupervised learning
Modeling
Marketing decision support
Fuzzy rules

ABSTRACT

The study here highlights the potential that analytical methods based on Knowledge Discovery in Databases (KDD) methodologies have to aid both the resolution of unstructured marketing/business problems and the process of scholarly knowledge discovery. The authors present and discuss the application of KDD in these situations prior to the presentation of an analytical method based on fuzzy logic and evolutionary algorithms, developed to analyze marketing databases and uncover relationships among variables. A detailed implementation on a pre-existing data set illustrates the method.

© 2012 Published by Elsevier Inc.

1. Introduction

Since the 1980s, increasing complexity, uncertainty, unpredictability, and ill-definition characterize competitive environments of companies worldwide (Hough & Duffy, 1987; Hunt, 2000). To be competitive, firms need now to develop adequate competences to effectively compete in a turbulent context (Grant, 2003; Lei, Hitt, & Bettis, 1996). These critical competencies include having the capability to process huge amounts of data, and generating and disseminating pertinent knowledge to aid managers' decision making (Sher & Lee, 2004). Notwithstanding, most firms' research and analytical methods are unable to cope with the informational needs of these new competitive environments (Tay & Lusch, 2005). Organizations now demand research and analytical methods that provide added-value information to support strategic and operational decisions (Avison, Eardley, & Powel, 1998). A widely accepted term, integrating the diversity of tools and computerized systems used by firms for this purpose, is marketing management support systems (MkMSS, see Van Bruggen & Wierenga, 2001; Wierenga & Van Bruggen, 1993, 1997, 2000).

The evolution of MkMSS parallels an evolution in the type of decision problems managers must address. Specifically, while some marketing problems are well-structured, many others, especially those belonging to the strategic sphere, are complex, ill-defined, or unstructured (Wierenga, 2010). Such unstructured problems contain elements, and/or relationships between elements, that are not known *a priori*. These problems are novel by definition, and therefore managers cannot depend on a pre-programmed decision protocol to solve such problems (Wierenga & Van Bruggen, 1997, 2000). For instance, a sales manager could receive a large number of customer complaints reporting unethical salesperson behavior. The novelty of a problem of this kind needs the manager to take an open approach to understanding the possible explanatory factors, and to subsequently develop a unique solution. Ill-defined problems such as this frequently require systems able to both work with and provide outputs expressed in qualitative terms, because managers must often use qualitative judgmental processes in these cases. Such systems form a specific type of MkMSS category (i.e., knowledge-driven), and Knowledge Discovery in Databases (KDD) methodologies are increasingly employed within knowledge-driven MkMSS. However, in spite of their potential in supporting marketing decisions, traditional quantitative data-driven MkMSS have not employed KDD methods (Shaw, Subramaniam, Tan, & Welge, 2001).

This study introduces an ad-hoc KDD method called Fuzzy-CSar. Fuzzy-Csar is a genetic fuzzy system, based on the combination of fuzzy logic and a genetic algorithm (see Zadeh, 1994). Fuzzy-CSar is an evolutionary unsupervised learning method, and aims at uncovering interesting and reliable information patterns (termed fuzzy association rules) in marketing databases. Fuzzy-CSar is able to work without *a priori* information on any relationships among the variables to hand. So, the search process is not driven by a relational structure

* Corresponding author.
E-mail addresses: aorriols@salle.url.edu (A. Orriols-Puig), fjmlopez@ugr.es (F.J. Martínez-López), casillas@decsai.ugr.es (J. Casillas), n.j.lee@aston.ac.uk (N. Lee).

0148-2963/$ – see front matter © 2012 Published by Elsevier Inc.
doi:10.1016/j.jbusres.2012.02.033

of reference (e.g. a model), and this characteristic provides clear benefits when Fuzzy-Csar is applied to new, non-habitual decisional scenarios.

Fuzzy-Csar's combination of fuzzy logic and genetic algorithms has a set of powerful advantages to users who must analyze novel scenarios. First, the use of fuzzy rules allows users (e.g. managers or scholars) working with linguistic semantics to use a convenient transformation of the original measurement scales to define the variables of the database. This particular aspect of the method allows a linguistic interpretation of the relationships among the variables, which is of clear benefit when using qualitative information.

A second benefit is that the kind of output offered by Fuzzy-CSar is similar to the internal rules in human reasoning. Such output is impossible without the use of fuzzy logic, and is a unique contribution of the Fuzzy-CSar method. Fuzzy-Csar's use of genetic algorithms (GAs) is also a key advantage of the method. GAs show better performance than alternative techniques in deriving the fuzzy models used in Fuzzy-CSar (Casillas & Martínez-López, 2009). Furthermore, the evolutionary nature of GAs facilitates decision-making in environments that are subject to change (Butel & Watkins, 2000).

Section 2 presents the background and discusses the managerial and academic utilities of using machine learning in marketing domains. Section 3 describes the main technical characteristics of the proposed method. Section 4 presents the study methodology and Section 5 analyzes the results. Finally, Section 6 concludes the work.

2. Utilities of KDD bottom–up approaches

2.1. Managerial purposes: aiding decision making

Both marketing scholars and practitioners make extensive use of theoretical models (i.e., *a priori* structures) to structure the analysis of databases. However, the use of models to drive knowledge search has limitations in certain decision-making situations. In particular, practicing managers may disregard pre-existing theoretical models in the belief that general relational structures are ineffective to solve their contextually-specific decisions. This situation is unlike the academic arena, where analytical processes are heavily conditioned by theoretical frameworks, research methodologies, and procedures, in accordance with the scientific method. As noted above, managers must face problems that are not properly structured or delimitated, and/or where no *a priori* information exists to orient analyses in databases.

In unstructured cases, analytical tools predicated on assumptions of routine, well-programmed problems are not adequate. Unstructured decision problems demand more creative solutions (Wierenga & Van Bruggen, 2000), and managers must be open-minded when assessing the application of new methods. In this regard, properly-designed artificial intelligence/knowledge-based methods have considerable potential to achieve good performance in unstructured, ill-defined marketing problems. Despite this potential, artificial intelligence/knowledge-based methods are underused at present to support decisions in marketing management, where data-driven, quantitative-oriented methods are dominant (Wierenga, 2010).

When applying a KDD process, two main approaches can be distinguished for what is termed the 'machine learning' stage (Zhang, 2004). The first approach is termed top–down, where the analyst uses an *a priori* relational structure for the set of variables defining the problem. The second approach is termed bottom–up, where the analyst does not use existing information on what structures may connect the data. The bottom–up approach is based on methods that automatically explore the data in order to uncover subjacent information patterns (i.e., connections between variables). Bottom–up methods have a different philosophy of application to that applied by most machine learning research, in which researchers usually set and test hypotheses by means of a top–down learning technique (Michael & Paul, 1999).

At first sight, the outputs provided by a bottom–up KDD process might seem more difficult to work with. Specifically, one may expect that managers would need to discard a lot of unexpected or unusable relationship patterns. However, the undefined and ill-structured nature of the problems of interest makes the application of model-driven analytical methods more difficult. Managers may be helped in making better decisions by the information patterns extracted from a database after an exploratory, unsupervised machine learning method. Bottom–up KDD methods can also stimulate the creativity of users, and motivate more divergent reasoning (see for example the ORAC models in Wierenga & Van Bruggen, 1997).

2.2. Academic purposes: the role of bottom–up methods for scientific knowledge generation

KDD-based methods at present seem to demonstrate their greatest potential when applied in the practitioner arena. However, KDD-based methods might be also useful for scientific research. Such methods can act as complementary analytical tools to discover knowledge. It is out of the current research scope to enter into debates on the nature of science and knowledge. Indeed, multiple monographs that treat this question in detail exist (e.g., Benton & Craib, 2001; Bishop, 2007; Chalmers, 1999). Nevertheless, one may make some overall observations regarding KDD methods and their plausibility as scientific research tools.

Certainly, the purely bottom–up use of machine learning tools does not match the hypothetico-deductive premises of the logic of justification in marketing theory (see Hunt, 2002a, 2002b). However, bottom–up methods are completely coherent with the logic of discovery. In this regard, the results provided by unsupervised machine learning methods are useful when following induction or, even better, abduction as scientific research procedures. Scholars highlight that, marketing researchers should also consider the utility of research applications outside the *a priori* theory-oriented when promoting creativity and progress in marketing knowledge development (e.g. Deshpandé, 1983; Peter & Olson, 1983). With this focus, marketing scholars might use KDD methods without the support of theories to guide the knowledge discovery process. However, Hunt's (2002a) reflection that one may treat these pieces of information as knowledge claims rather than scientific knowledge is reasonable (see also Lee & Lings, 2008). Even so, the machine learning approach has obvious applications in exploratory marketing research, to formulate hypotheses for later testing (Popper, 1959).

3. Description of the Fuzzy-CSar method

The Fuzzy-CSar method automatically evolves a population of association rules that denote strong and relevant relationships among the variables of a given problem. This section briefly introduces the Fuzzy-CSar method, and describes the learning procedure. The section places special emphasis on detailing the knowledge representation, and also on the use of multi-item fuzzification to deal with the particularities of marketing data.

3.1. Knowledge representation

Fuzzy-CSar evolves a *population* of association rules, which are usually referred to as *classifiers*. At the end of the learning process, the population is expected to capture the strongest and most relevant associations among the variables. The user sets the maximum size of the population. This maximum fixes an upper bound on the number of relevant associations that can be found; that is, at maximum, the system will be able to discover as many relevant relationships as the 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: ***if*** x_i *is* A_i *and*

… *and* x_j *is* A_j ***then*** x_c *is* A_c, where the antecedent contains ℓ_a input variables x_i, …, x_j ($0<\ell_a<\ell$, where ℓ is the number of variables of the problem) and the consequence consists of a single variable x_c which is not present in the antecedent. In this study's presentation, each variable is represented by a *linguistic term* or *label*, A_i, which can be defined by the user. This structure allows a number of variables to be antecedents, but requires a single variable as the consequence. With this strategy, the researcher is searching for sets of variables with certain values that cause another variable to occur. Such rules can therefore be interpreted as a causal relationship between certain values of the variables in the antecedent(s) and certain values of the consequent variable.

In addition to the association rule itself, each classifier has two main parameters; support and confidence. Support is an indicator of the frequency of occurrence of the rule, and confidence denotes the strength of the association. The Fuzzy-CSar system is designed to search for the rules with high support (i.e., rules that denote a relationship that can be found frequently), and high confidence (i.e., rules in which the values of the variables in the antecedent determine the value of the variable in the consequent).

3.2. Multi-item fuzzification

Marketing and social science research commonly uses multi-item measurement scales. Fuzzy-CSar uses multi-item fuzzification to cope with multi-item measurement (Martínez-López & Casillas, 2009). Multi-item fuzzification is based on the premise that each individual item provides partial information about the corresponding construct (i.e., an unobserved or first-order variable). As such, the matching degree can be computed as the aggregation of the information given by each item. The matching degree of a variable *i* with the vector of items $x_i=(x^i_1, x^i_2, \ldots, x^i_{pi})$ is

$$\mu_{A_i}(x_i)= \max_{h_i=1}^{p_i}\mu_{A_i}(x^i_{p_i}),$$

where $\mu_{Ai}(x^i{}_{pi})$ is the matching degree of the variable i represented by the linguistic term A_i with the input $x^i{}_{pi}$. This study uses the maximum as the union operator, implemented as a sum bounded to 1. However, Fuzzy-CSar is also able to work with single-item variables and measures.

3.3. Learning process

Fuzzy-CSar follows an online learning scheme to evolve a population of highly relevant rules, starting with an empty population and learning as new training examples are sampled. More specifically, the system receives a new training example at each learning iteration, and then takes a number of 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] does not contain enough classifiers, the *covering operator* is triggered to create new classifiers. Then, classifiers in [M] are organized into *association set candidates*.

Each association set candidate is given a probability of selection 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 that aims at diminishing the number of rules that express similar associations among variables. Then, the parameters of all the classifiers in [M] are updated according to the information provided by the current example. At the end of the iteration, a genetic algorithm is applied to [A] to discover new promising rules. This process is repeatedly applied to update the parameters of existing classifiers and create new promising rules.

4. Problem description and methodology

4.1. Database and marketing problem used for the experimental application

This study demonstrates Fuzzy-CSar by applying the method to an existing database. Cadogan, Lee, Tarkiainen, and Sundqvist (2009) collect data from 154 Finnish sales managers via mail survey and apply structural equation modeling in order to test hypotheses regarding salespeople's unethical behavior. Cadogan et al.'s (2009) model is based on theories describing specific antecedent factors related to the organizational environment, grouped into *sales manager-specific factors* and *sales team-specific factors*. Regarding the manager-specific factors, Cadogan et al. (2009) consider the following four variables: *behavior control* (H2: +), *outcome control* (H3: −), *relativism* (H4: −), and *idealism* (H5: +). With regard to the sales team-specific factors, Cadogan et al. (2009) consider *job insecurity* (H6: −), *cooperation* (H7: +), and *tactical performance* (H8: +). The underlying argument by Cadogan et al. (2009) is that both sales manager- and sales team-specific factors influence the *ethical standards* of the sales team (see the sign of the hypothesized relationship in brackets), which in turn influence *ethical sales behavior*. Cadogan et al. (2009) find no statistical support for hypotheses H3, H5 and H8, but do find statistical support for a positive relationship between the variables in hypotheses H2 and H7. Finally, Cadogan et al. (2009) report statistical support for a negative relationship between the variables in hypotheses H1, H4, and H6.

Cadogan et al.'s (2009) approach is clearly theory-driven, and is conditioned by an *a priori* relational structure of the database. As these hypotheses specify only a limited number of relationships between variables, some key relationships may not have been discovered. Of course, researchers using SEM can use either modification indices or Lagrange multipliers (dependent on the specific SEM package used) to examine possible unhypothesized relationships between variables. However, even if Cadogan et al. (2009) use these techniques, they remain able to only consider relations between pairs of variables, and between the full ranges of each variable. The possibility exists of relevant correlations between more than two variables, or correlations only at specific ranges of certain variables. While one may test such hypotheses in an SEM framework, the complexity of doing so makes is unfeasible to do so without *a priori* hypotheses. Thus, in order to extend this analysis, this study uses Fuzzy-CSar to identify relationships not captured by Cadogan et al. (2009). As such the Fuzzy-CSar approach complements, rather than replaces, the classical theory-driven process of knowledge generation.

4.2. Experimental methodology

To demonstrate the usefulness of Fuzzy-CSar, this study performs a set of experiments aimed at studying (1) the robustness of Fuzzy-CSar in dealing with marketing data, (2) the ability of the system to identify and support the original hypotheses, and (3) the ability of the system to discover new relevant relationships between variables.

Experiment 1 The goal of the first experiment is to determine whether the system can evolve association rules that indicate any causal relationship between any variable and the variables *ethical standards* and *unethical behavior*. Therefore, system follows a theory-driven (top–down) approach, by forcing that the rule's consequent can only be either ethical standards or unethical behavior. This experimental design matches the typical academic approach (Section 2.2).

Experiment 2 The second experiment investigates whether Fuzzy-CSar can discover new relationships between any variables. For this purpose, system runs without any

requirement on the specific variables that are the antecedent and consequent of the rule. Thence, with this configuration the system has the freedom to create rules that denote any causal relationship within the variables of the data set. Such a design is typical of managerial applications (see Section 2.1).

For both experiments, the system is configured with a population size of 6.400 rules and the following parameters: $P_{\#}=0.5$, $P_{\chi}=0.8$, $\{P_{I/B}, P_{\mu}, P_C\}=0.1$, $_\theta_{GA}=50$, $_\theta_{exp}=1\ 000$, $conf_0=0.95$, $\nu=1$, $\delta=0.1$. In addition, all the variables use Ruspini's strong fuzzy semantics with three linguistic terms: *small, medium*, and *large*.

5. Analysis of the results

In the first instance, the authors consider only the rules that lay in the Pareto set, that is, those rules for which no other rule with both better support and confidence exists. Following this, the authors compare the quality of the rules obtained by the two configurations of Fuzzy-CSar. Subsequently, the authors analyze whether the created rules support the hypotheses formulated by Cadogan et al. (2009). Finally, the authors extend the analysis to examine the new rules evolved by Fuzzy-CSar, and highlight those that show new and potentially interesting causal relationships.

5.1. Analysis of the quality of the rules in the objective space

The first objective of the study is to analyze the significance of the rules in the objective space, and thus to show the abilities of Fuzzy-CSar to discover promising rules. To this end, Fig. 1 shows the Pareto set obtained with the different configurations, that is, the support and the confidence of the most relevant rules evolved by the system.

The curves *Res-ES* and *Res-UB* show the confidence and support of the most relevant rules (from the objective point of view) found by Fuzzy-CSar when the system is configured to force the variables *ethical standards* and *unethical behavior* respectively to be in the consequent. Fig. 1 illustrates that the system can find rules which denote a causal relationship between input variables and *unethical behavior* with high support and confidence (curve *Res-UB*). On the other hand, the system can also find rules that denote strong causal relationships between the input variables and *ethical standards* (curve *Res-ES*), although the confidence and support of these rules is lower than those for *ethical behavior*. The lower relevance of the *ethical behavior* rules may indicate that the input variables describe *unethical behavior* more accurately than they describe *ethical standards*.

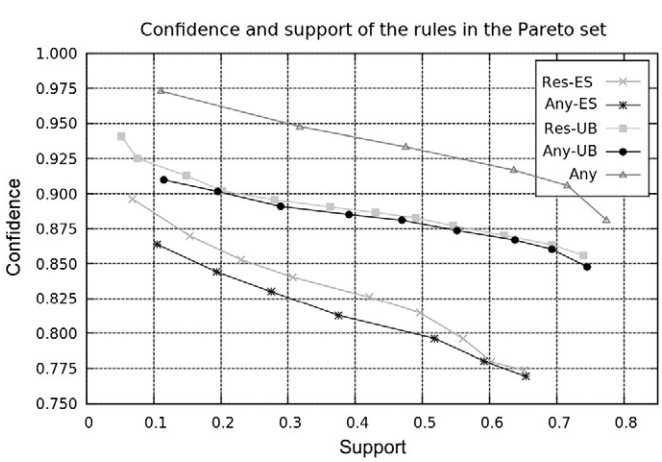

Fig. 1. Average Pareto front obtained by Fuzzy-CSar with different runs. Res-ES rand Res-UB show the Pareto set of the rules that has *ethical standards* and *unethical behavior* for the runs in which these two variables are forced to be in the consequent (experiment 1). Any-ES and Any-UB show the same information but for the runs in which any variable can be in the consequent (experiment 2). Finally, Any reports the Pareto set considering all the rules regardless of the variables they have in the antecedent.

The experiments regarding *Res-ES* and *Res-UB* serve to set a baseline of the interestingness of the rules that the system could find, subject to the restriction that the rules can only contain either *ethical standards* or *ethical behavior* in their consequent. Nevertheless, one of the main advantages of Fuzzy-CSar is that the system can discover rules that denote a causal relationship of the input variables with any variable, not just *ethical standards* and *ethical behavior*. The authors demonstrate this with the second experiment, where Fuzzy-CSar is not restricted to finding only rules with specified consequents. The results are shown by the *Any-ES*, *Any-UB* and *Any* curves shown on Fig. 1. The three curves show the most relevant rules with (1) *ethical standards*, (2) *unethical behavior*, and (3) any of the nine variables in the consequent of the rule respectively.

The results of experiment two highlight the robustness and added value of Fuzzy-CSar. First, the curves *Any-ES* and, especially, *Any-UB* are very close to their corresponding curves in the first experiment, that is, *Res-ES* and *Res-UB*. This closeness highlights the robustness of Fuzzy-CSar. Specifically, Fuzzy-CSar is able to find the most relevant rules that denote a causal relationship with the variables *ethical standards* and *unethical behavior* even though the search space has been hugely increased in this second experiment. Second, the curve *Any* shows that the system can find rules with more interest by considering any variable in the consequent of the rules. This result emphasizes the existence of strong relationships between variables other than *ethical standards* and *unethical behavior* that have potential for further study. In the following subsections, the authors conduct a more detailed study of the information that these rules actually contain.

5.2. Does Fuzzy-CSar Support the Original Model Hypotheses?

The authors now move to studying the new information that the rules evolved by Fuzzy-CSar can report to the expert analyst. This section first studies whether the rules obtained by Fuzzy-CSar support and/or extend the theory-driven conclusions drawn in Cadogan et al. (2009). This focus is interesting to academic researchers, but also to managers interested in applying a top–down approach with Fuzzy-CSar.

Cadogan et al.'s (2009) use SEM to test their model, which is popular in academic management research (Shook, Ketchen, Hult, & Kacmar, 2004), because of the ability of SEM to cope with complex models with latent constructs. As such, even though scholars argue that SEM has less use as a decision support tool in managerial practice (see Laurent, 2000; Steenkamp & Baumgartner, 2000), this study focuses on the SEM results as a benchmark to assess the utility of Fuzzy-CSar. However, the intention of the demonstration is to show the complementarities of Fuzzy-CSar to SEM, not to imply the superiority of Fuzzy-CSar over SEM. Further, while space constraints preclude a more detailed description, the study does synthesize several incremental benefits of Fuzzy-CSar.

The authors check whether Fuzzy-CSar supports the hypotheses of Cadogan et al. (2009), by searching the final populations of rules for rules with a) high support and confidence, and b) only one variable in the antecedent and one in the consequent. The authors only consider rules which correspond to either positive or negative causal relationships, such rules have either the same or completely opposite linguistic terms respectively. More specifically, considering the three linguistic terms employed – i.e., *small* (S), *medium* (M), and *large* (L) – a positive rule is one such as *if* var_i *is* S *then* var_j *is* S, while a negative rule is one such as *if* var_i *is* S *then* var_j *is* L.

The analysis returns rules relating to the following hypotheses of Cadogan et al. (2009), where *s* and *c* stand for support and confidence respectively:

1. *High unethical behavior low ethical standards* with $s=0.653$ and $c=0.778$. This rule supports hypothesis 1 in Cadogan et al. (2009) indicating that people with high unethical behavior have low ethical standards.
2. *Low behavioral control low ethical standards* with $s=0.580$ and $c=0.780$. This rule supports hypothesis 2 in Cadogan et al. (2009) indicating that the lower the extent to which sales managers' evaluations are based on the input behavior of their sales teams, the lower the ethical standards of these sales teams.
3. *Low outcome control low ethical standards* with $s=0.651$ and $c=0.779$. This rule supports the idea that managers evaluated on low outcome control rate their teams' ethical standards lower. Although this hypothesis is not supported by Cadogan et al. (2009), Fuzzy-CSar highlights the strength of the relationship between the two variables at low values. Fuzzy-CSar also does not detect a strong relationship between variables when both take high values. Thus, Fuzzy-CSar provides more precise information than SEM about this relationship.
4. *High relativism low ethical standards* with $s=0.565$ and $c=0.810$. This rule supports hypothesis 4 in Cadogan et al. (2009), upholding the belief that managers with high relativism promote a situational view or the rightness or wrongness in their teams.
5. *Low idealism high ethical standards* with $s=0.186$ and $c=0.234$. This rule supports the hypothesis that managers' levels of idealism are negatively related to sales team ethical standards. Nonetheless, both the support and the confidence of this rule are very low. This is the only rule that Fuzzy-CSar finds for hypothesis 5, and thus the results of the experiment tend to agree with the results of Cadogan et al. (2009) who find no support for this hypothesis.
6. *High job insecurity low ethical standards* with $s=0.330$ and $c=0.819$. This rule supports hypothesis 6 in Cadogan et al. (2009) by upholding that higher levels of job insecurity in the sales team are associated with lower ethical standards.
7. *Low team cooperation low ethical standards* with $s=0.544$ and $c=0.799$. This rule supports hypothesis 7 in Cadogan et al. (2009), indicating that low levels of information sharing are associated with low ethical standards.
8. No association rule was created to support hypothesis 8, supporting the conclusion of Cadogan et al. (2009) who do not find a significant relationship between these tactical performance and ethical standards.

The Fuzzy-CSar results not only confirm the conclusions extracted in Cadogan et al. (2009), but also provide more detail on the relationships than SEM is able to provide. The ability of Fuzzy-CSar to identify relationships between specific ranges of variables' values allow the authors to model Cadogan et al.'s (2009) hypotheses more precisely, and identify some relationships that need to be reconsidered (e.g. hypothesis 3).

The results therefore show how Fuzzy-CSar can, when used with a top–down approach, enrich results originally provided by SEM. The additional information provided by Fuzzy-CSar is useful whether applied with an academic or managerial orientation, helping to overcome some of the shortcomings of SEM. Fuzzy-CSar a) does not assume any specific kind of relation (e.g. linear, quadratic, etc.) among variables in the database, b) allows the analysis and interpretation of relationships among several variables, and c) allows relationships with different degrees of intensity to be uncovered. Fuzzy-CSar, used with a KDD top–down approach, is therefore more flexible and capable of handling qualitative information, than the statistical estimation methods habitually applied in marketing (see, also: Gatignon, 2000; Van Bruggen & Wierenga, 2000)

5.3. Can Fuzzy-CSar help to discover new/unexpected relationships?

In addition to the added value provided by Fuzzy-CSar in the analysis above, Fuzzy-CSar also enables the discovery of relationships that Cadogan et al. (2009) do not consider. To demonstrate the potential of Fuzzy-CSar in this regard, the following section provides interesting examples of such rules. However, the authors stress that such emergent relations must be interpreted in light of conceptual logic and theory, rather than simply accepted at face value.

1. *Low outcome control high unethical behavior* with $s=0.720$ and $c=0.859$. This rule provides reliable information on a direct relation between these two constructs. Fuzzy-CSar analysis thus suggests that the ethical standards of the sales team may not be necessary to mediate this relation at these more extreme values (i.e., low for the former and high for the latter).
2. *Low relativism low idealism* with $s=0.658$ and $c=0.932$. This rule shows a relation between two constructs not considered in Cadogan et al.'s (2009) theoretical model. Moreover, the rule is an unexpected scenario that differs from what may be theoretically reasoned. Such a rule needs additional investigation.
3. *High relativism* and *low cooperation low idealism* with $s=0.466$ and $c=0.940$. The final rule evolved by Fuzzy-Csar is a curious rule that relates sales manager and sales team characteristics. In particular, when a sales manager with a high sense of relativism manages a sales team with a low tendency to cooperate, the manager perhaps reduces his/her idealistic beliefs to low levels.

6. Conclusions

This paper demonstrates the potential of machine learning techniques to help in the discovery of interesting, useful and novel information in marketing problems. More specifically, the authors propose the use of Fuzzy-CSar, a general-purpose unsupervised learning technique that evolves a set of association rules online, and that also uses adapted inference mechanisms to deal with the particularities of marketing data. Fuzzy-CSar does not need *a priori* hypotheses, but instead automatically discovers the strongest and most frequent associations between variables. Fuzzy-CSar also enables the analyst to discount potentially spurious relationships, which are represented in the form of rules with very low support.

The experimental analysis in this study emphasizes the added value provided by the application of machine learning techniques to data-analytic problems. The authors show in the first instance that Fuzzy-CSar supports the conclusions drawn in the original analysis by Cadogan et al. (2009). Fuzzy-CSar also uncovers additional information that a) suggests reconsideration of some of Cadogan et al.'s (2009) initial hypotheses and b) illustrates novel relationships among two or more variables. The successful results of these experiments should encourage further investigation and use of machine learning techniques, as a complement to existing business research analysis methods.

Acknowledgments

The authors thank the Ministerio de Educación y Ciencia for its support under projects TIN2008-06681-CO6-01 and TIN2008-06681-CO6-05, Generalitat de Catalunya for its support under grant 2005SGR-00302, and Junta de Andalucía for its support under project P07-TIC-3185. Further, the authors thank John Cadogan, Anssi Tarkiainen, and Sanna Sundqvist for their kind consent to the use of their database.

References

Avison, D. E., Eardley, W. A., & Powel, P. (1998). Suggestions for capturing corporate vision in strategic information systems. *Omega*, *26*(4), 443–459.

Benton, T., & Craib, I. (2001). *Philosophy of Social Science: Philosophical Issues in Social Thought (Traditions in Social Theory)*, Palgrave Macmillan.

Bishop, R. (2007). *The Philosophy of the Social Sciences: An Introduction.* Continuum.

Butel, L., & Watkins, A. (2000). Evolving complex organizational structures in new and unpredictable environments. *Journal of Business Research*, *47*, 27–33.

Cadogan, J. W., Lee, N., Tarkiainen, A., & Sundqvist, S. (2009). Sales manager and sales team determinants of salesperson ethical behavior. *European Journal of Marketing*, *43*(7/8), 907–937.

Casillas, J., & Martínez-López, F. J. (2009). Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. *Expert Systems with Applications*, *36*(2, part 1), 1645–1659.

Chalmers, A. F. (1999). *What Is This Thing Called Science: An Assessment of the Nature and Status of Science and Its Methods* (3rd Ed.). Open University Press.

Deshpandé, R. (1983). Paradigms lost: on theory and method in research in marketing. *Journal of Marketing*, *47*(4), 101–110.

Gatignon, H. (2000). Commentary on Peter Leeflang and Dick Wittink's Building models form marketing decisions: past, present and future. *International Journal of Research in Marketing*, *17*. (pp. 209–214) (Special issue on marketing modelling on the threshold of the 21st century).

Grant, R. M. (2003). Strategic planning in a turbulent environment: Evidence from the oil majors. *Strategic Management Journal*, *24*(6), 491–517.

Hough, P. K., & Duffy, N. M. (1987). Top management perspectives on decision support systems. *Information & Management*, *12*(1), 21–30.

Hunt, S. D. (2000). *A general theory of competition.* Thousand Oaks (CA): Sage Publications.

Hunt, S. D. (2002). *Foundations of Marketing Theory: Toward a general theory of marketing.* M.E. Sharpe, Inc.

Hunt, S. D. (2002). *Controversy in Marketing Theory: For reason, realism, truth, and objectivity.* M.E. Sharpe, Inc.

Lee, N., & Lings, I. (2008). *Doing Business Research.* London: Sage.

Lei, D., Hitt, M. A., & Bettis, R. (1996). Dynamic core competences through meta-learning and strategic context. *Journal of Management*, *22*(4), 549–569.

Laurent, G. (2000). Improving the external validity of marketing models: A plea for more qualitative input. International Journal of Research in Marketing, Vol. 17. *Special issue on "Marketing modelling on the threshold of the 21st century"* 177–182.

Martínez-López, F. J., & Casillas, J. (2009). Marketing Intelligent Systems for consumer behaviour modelling by a descriptive induction approach based on Genetic Fuzzy Systems. *Industrial Marketing Management*, *38*(7), 714–731.

Michael, C. H., & Paul, G. (1999). Special Section: data mining. *Journal of Management Information Systems*, *16*(1), 11–16.

Peter, J. P., & Olson, J. C. (1983). Is science marketing? *Journal of Marketing*, *47*(4), 111–125.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, *31*, 127–137.

Sher, P. J., & Lee, V. C. (2004). Information technology as a facilitator for enhancing dynamic capabilities through knowledge management. *Information & Management*, *41*(8), 933–945.

Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An assessment of the use of structural equation modelling in strategic management research. *Strategic Management Journal*, *25*, 397–404.

Steenkamp, J. -B. E. M., & Baumgartner, H. (2000). On the use of structural equation models for marketing modeling. *International Journal of Research in Marketing*, *17*, 195–202 (Special issue on marketing modelling on the threshold of the 21st century).

Tay, N. S. P., & Lusch, R. F. (2005). A preliminary test of Hunt's General Theory of Competition: Using artificial adaptive agents to study complex and ill-defined environments. *Journal of Business Research*, *58*, 1155–1168.

Van Bruggen, G. H., & Wierenga, B. (2000). Broadening the perspective on marketing decision models. *International Journal of Research in Marketing*, *17*, 159–168 (Special issue on marketing modelling on the threshold of the 21st century).

Van Bruggen, G. H., & Wierenga, B. (2001). Matching management support systems and managerial problem-solving modes: The key to effective decision support. *European Management Journal*, *19*(3), 228–238.

Wierenga, B. (2010). Marketing and artificial intelligence: Great opportunities, reluctant partners. In J. Casillas, & F. J. Martínez-López (Eds.), *Marketing Intelligent Systems using Soft Computing: Managerial and Research Applications,* Springer, 1–8.

Wierenga, B., & Van Bruggen, G. T. (1993). Marketing Management Support Systems: Evolution, Adoption and Effectiveness. In ESOMAR/EMAC/AFM (Ed.), *Information Based Decision Making in Marketing: Reconciling the Needs and Interests of Decision Makers, Data Collectors and Data Providers* (pp. 1–20).

Wierenga, B., & Van Bruggen, G. T. (1997). The integration of marketing problem-solving modes and marketing management support systems. *Journal of Marketing*, *61*, 21–37.

Wierenga, B., & Van Bruggen, G. T. (2000). *Marketing management support systems: principles, tools and implementation,* Kluwer Academic Publishers.

Zadeh, L. A. (1994). Soft computing and fuzzy logic. *IEEE Software*, *11*, 48–56.

Zhang, M. (2004). Introduction to data mining and KDD process, Technical report, Victoria University of Wellington. Available at http://www.mcs.vuw.ac.nz/courses/COMP422/2004T2/Lecturenotes/week01.pdf