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.

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*