Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 36 (2009) 1645–1659

Expert Systems with Applications

www.elsevier.com/locate/eswa

Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling ☆

Jorge Casillas [a,*], Francisco J. Martínez-López [b]

[a] Department of Computer Science and Artificial Intelligence, University of Granada, E-18071 Granada, Spain
[b] Marketing Department, University of Granada, E-18071 Granada, Spain

Abstract

The main problem currently faced by market-oriented firms is not the availability of information (data), but the possession of appropriate levels of knowledge to take the right decisions. This is common background for firms. In this regard, marketing professionals and scholars highlight the necessity for knowing and explaining consumers' behaviour patterns in an increasingly efficient way. The use of new knowledge discovery methods, able to exploit such data, may represent a relevant source of competitive advantage.

In marketing, the information about most consumer variables of interest is usually obtained by means of questionnaires containing a diversity of items. It is also frequent that marketing modellers make use of unobserved variables to build the consumer models; i.e., abstract variables that need to be measured by means of a set of observed variables or items associated with them. In these cases, the value of a certain unobserved variable cannot be assigned to a number, but to a potentially scattered set of numbers. This fact disables the application of conventional data mining techniques to extract knowledge from them.

In this paper, we present a new approach that is able to deal with this kind of uncertain data by using a multiobjective genetic algorithm to derive fuzzy rules. Specifically, we propose a complete methodology that considers the different stages of knowledge discovery: data collection, data mining, and knowledge interpretation. This methodology is experimented on a consumer modelling application in interactive computer-mediated environments.

Keywords: Consumer behaviour; Fuzzy logic; Genetic algorithms; Knowledge extraction; Machine learning; Marketing

1. Introduction

1.1. Background

Nowadays, it is usual to find numerous academic and professional management-related contributions that show a clear tendency towards defining the current business environment of organizations as hypercompetitive (D'Aveni, 1994). Unlike several decades ago, when companies were mainly concerned with having enough information to guide their decisional processes, the problem faced by organizations has been become increasingly centred, from the middle 1980s till now, on the possession of high quality, thus most valued, and better information about their business framework than their competitors. This is a widely accepted idea that has driven the evolution of the management information systems (MIS), in terms of their design, use and support functions required, from those based on the data, to the ulterior knowledge-based systems whose first exponent were the expert systems (Casey & Murphy, 1994; Sisodia, 1992; Talvinen, 1995; Van Bruggen & Wierenga, 2000).

This question of the improvement of the quality of the systems used to manage the variety of information owned by firms becomes even more important when it is analyzed

☆ This work was sponsored in part by the Spanish Ministry of Science and Education under Grant No. TIN2005-08386-C05-01 and by the European Regional Development Fund.

* Corresponding author. Tel.: +34 958240804; fax: +34 958243317.
E-mail addresses: casillas@decsai.ugr.es (J. Casillas), fjmlopez@ugr.es (F.J. Martínez-López).

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.11.035

within the framework of the marketing function; the business area primarily responsible for managing the relations with consumers, i.e., the firm's target. As Li, Kinman, Duan, and Edwards (2000) note, considering the characteristics related to the competitive environment of firms, marketing strategies directed to markets must be based on an accurate knowledge of the consumers' preferences and behaviour. In this task, the application of suitable marketing management support systems (MkMSS) to the analysis of data plays a notable role (Wierenga & Van Bruggen, 1997, 2000). It is not unusual, therefore, to observe the intensification in the use of knowledge-based MkMSS by firms in recent years (Shim et al., 2002; Wedel, Kamakura, & Böckenholt, 2000).

This evolution of the MkMSS towards systems based on methodologies imported from the artificial intelligence area have tried to fulfil the demands of marketing managers and modellers in terms of working with methods of analysis that are more flexible, powerful and robust, and capable of providing greater and improved information with respect to consumers' behaviour (Lilien, Kotler, & Moorthy, 1992). In this sense, though it is well known that the marketing expert systems were, in line with the MIS framework, the first knowledge-based systems applied to support the marketing managers' decision processes, there have been significant and interesting advances, some of them very recent, such as those based on artificial neural networks, case-based reasoning, clustering, decision trees, or fuzzy systems (Akhter, Hobbs, & Maamar, 2005; Ha & Park, 1998; Li et al., 2000; Wierenga & Van Bruggen, 2000). In any case, regardless of the marketing knowledge-based system we consider, each has one thing in common, the use of knowledge discovery in databases (KDD) methodologies (Fayyad, Piatesky-Shapiro, Smyth, & Uthurusamy, 1996), and hence, data mining (machine learning) paradigms.

KDD implies the development of a process compounded by several stages that allow the conversion of low-level data into high-level knowledge, where the data mining is considered the core stage of such a process (Mitra, 2002). Nevertheless, it is important to be aware of the fact that the application of the data mining stage alone would be insufficient to undertake, with rigor and guarantees of success, a process of KDD (Fayyad & Simoudis, 1995).

It is well known that KDD may offer excellent results when applied to marketing databases in general, as well as to the analysis of the behaviour of consumers in particular, though the development and application of specific KDD methodologies to marketing problems is still incipient (Liao & Chen, 2004). In this regard, we believe that the benefits provided by KDD should not only motivate its use in firms that are clearly interested in improving the efficiency of their MkMSS, thus their marketing decision making, but also in marketing academics. Specifically, in our opinion, the academics' efforts must be focused on two main questions with regard to this: first, an intelligent, oriented and selective increase in the use of the KDD techniques, based on their properties for solving the marketing problem faced by the academics; and second, an active research to adapt generic KDD methodologies to the specific characteristics of the marketing problems to which they are going to be applied. In this sense, the main contribution of this paper focuses on the latter.

1.2. Scope of the paper

In KDD, we can distinguish between two different approaches (Lavrac, Cestnik, Gamberger, & Flach, 2004): *predictive* induction and *descriptive* induction. The difference lies in the main objective pursued and, therefore, the learning method used to attain it. On the one hand, predictive induction looks to generate legible models that describe with the highest reliability the data set that represents the analyzed system. In that case, the goal is to use the model obtained to simulate the system, thus reaching an explanation of its complex behaviour. On the other hand, descriptive induction looks for particular (interesting) patterns of the data set. In that case, we do not achieve a global view of the relationships among variables but we discover a set of rules (different enough among them) that are statistically significant.

This paper focuses on predictive induction to extract useful knowledge guided by theoretical (causal) models used in the discipline of consumer behaviour. In other words, the machine learning stage is driven by a set of relations among variables previously determined by the marketing expert. To do that, we develop a complete KDD methodology, adapted to the kind of causal structures, variables and measurement models usually used in consumer behaviour modelling. Hence, we reflect on and give specific solutions, adapted to the marketing problem we face, to the variety of questions associated with every stage of the KDD process; i.e., pre-processing, machine learning and post-processing. Basically, the benefits we provide with this methodology try to cover the academic as well as the professional fields, though we mainly highlight the interesting qualities of its practical applicability in order to help marketing managers to better predict consumer behaviour. Specifically, association fuzzy rules, with input and output variables previously fixed by the theoretic model of reference, are used. The extraction is performed by means of genetic fuzzy systems (Cordón, Herrera, Hoffmann, & Magdalena, 2001), i.e., genetic algorithms (GAs) used to learn fuzzy rules. Two questions arise at this stage: *why fuzzy rules?* and *why GAs*?

The use of fuzzy rules (instead of other knowledge representations such as interval rules, decision trees, support vectors, neural networks and) is justified mainly by the kind of data set we are dealing with (see Section 3.1). In our case, each variable is composed of a set of parameters (items) that add uncertainty to the data, since each provides partial information to describe the variable. Moreover, we are able to transform with ease the available