Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 182 (2007) 340–355
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Decision Support

Rough set-based logics for multicriteria decision analysis

Tuan-Fang Fan [a,b,*], Duen-Ren Liu [a], Gwo-Hshiung Tzeng [c,d]

[a] *Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan*
[b] *Department of Computer Science and Information Engineering, National Penghu University, Penghu 880, Taiwan*
[c] *Graduate Institute of Management of Technology, National Chiao Tung University, Hsinchu 300, Taiwan*
[d] *Kainan University, Taoyuan 338, Taiwan*

Received 14 December 2004; accepted 17 August 2006
Available online 13 October 2006

Abstract

In this paper, we propose some decision logic languages for rule representation in rough set-based multicriteria analysis. The semantic models of these logics are data tables, each of which is comprised of a finite set of objects described by a finite set of criteria/attributes. The domains of the criteria may have ordinal properties expressing preference scales, while the domains of the attributes may not. The validity, support, and confidence of a rule are defined via its satisfaction in the data table.

Keywords: Rough sets; Fuzzy sets; Decision logic; Multicriteria decision analysis; Artificial intelligence

1. Introduction

The theory of knowledge has long been an important topic in many academic disciplines, such as philosophy, psychology, economics, and artificial intelligence, whereas the storage and retrieval of data is the main concern of information science. In modern experimental science, knowledge is usually acquired from observed data, which is a valuable resource for researchers and decision-makers. However, when the amount of data is large, it is difficult to analyze the data and extract knowledge from it. With the aid of computers, the vast amount of data stored in relational data tables can be transformed into symbolic knowledge automatically. Thus, intelligent data analysis has received a great deal of attention in recent years.

While data mining research concentrates on the design of efficient algorithms for extracting knowledge from data, how to bridge the semantic gap between structured data and human-comprehensible concepts has been a long-lasting challenge for the research community. Kruse et al. (1999) called this the *interpretability*

* Corresponding author. Address: Institute of Information Management, National Chiao-Tung University, Hsinchu 300, Taiwan.
E-mail addresses: dffan@npu.edu.tw, tffan.iim92g@nctu.edu.tw (T.-F. Fan), dliu@iim.nctu.edu.tw (D.-R. Liu), ghtzeng@cc.nctu.edu.tw (G.-H. Tzeng).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.08.029

problem of intelligent data analysis. Since discovered knowledge is only useful for a human user when he can understand its meaning, the knowledge representation formalism plays an important role in the utilization of the induced rules. A good representation formalism should have clear semantics so that a rule can be effectively validated with respect to the given data tables. In this regard, logic is one of the best choices. As noted by Zadeh (1996), humans usually compute with words instead of numbers, so if we can incorporate linguistically meaningful terms into the representation formalism, the induced rules may be more useful to human decision-makers.

The rough set theory proposed by Pawlak (1982) provides an effective tool for extracting knowledge from data tables. To represent and reason about the extracted knowledge, a decision logic (DL) was proposed in Pawlak (1991). The semantics of the logic is defined in a Tarskian style through the notions of models and satisfaction. While DL can be considered an example of classical logic in the context of data tables, different generalizations of DL corresponding to some non-classical logics are also desirable from the viewpoint of knowledge representation. For example, to deal with uncertain or incomplete information, some generalized decision logics have been proposed (Fan et al., 2001; Liau and Liu, 1999, 2001; Yao and Liau, 2002; Yao and Liu, 1999).

When rough set theory is applied to multi-criteria decision analysis (MCDA), it is crucial that preference-ordered attribute domains and decision classes be dealt with (Greco et al., 1997, 1998, 1999a, 2000, 2001a, 2002, 2004; Slowinski et al., 2002b). The original rough set theory cannot handle inconsistencies arising from violations of the dominance principle due to its use of the indiscernibility relation. Therefore, in the above-mentioned works, the indiscernibility relation is replaced by a dominance relation to solve the multi-criteria sorting problem, and the data table is replaced by a pairwise comparison table to solve multi-criteria choice and ranking problems. The approach is called the dominance-based rough set approach (DRSA). For MCDA problems, DRSA can induce a set of decision rules from sample decisions provided by decision-makers. The induced rules form a comprehensive preference model and can provide recommendations about a new decision-making environment.

A strong assumption about data tables is that each object takes exactly one value with respect to an attribute. However, in practice, we may only have incomplete information about the values of an object's attributes. Thus, more general data tables and decision logics are needed to represent and reason about incomplete information. For example, set-valued and interval set-valued data tables have been introduced to represent incomplete information (Kryszkiewicz, 1998; Kryszkiewicz and Rybiński, 1996a,b; Lipski, 1981; Yao and Liu, 1999). A generalized decision logic based on interval set-valued data tables is also proposed in Yao and Liu (1999). In these formalisms, the attribute values of an object may be a subset or an interval set in the domain. Since crisp subsets and interval sets are both special cases of fuzzy sets, further generalization of data tables is desirable to represent uncertain information. In data tables containing such information, an object can take a fuzzy subset of values for each attribute. To represent knowledge induced from uncertain data tables, the decision logic also needs to be generalized.

DRSA has also been extended to deal with missing values in MCDA problems (Greco et al., 2001a; Slowinski et al., 2002b). A data table with missing values is a special case of uncertain data tables. Therefore, we propose further extending DRSA to uncertain data tables and fuzzy data tables. In this paper, we present a logical treatment of DRSA in precise data tables, as well as uncertain and fuzzy data tables. Our approach is concerned with variants of DL for data tables.

The remainder of the paper is organized as follows. In Section 2, we review the decision logic proposed by Pawlak. In Sections 3–6, we respectively present generalized DL for preference-ordered data tables, preference-ordered uncertain data tables, preference-ordered fuzzy data tables, and pairwise comparison tables. For each logic, the syntax and semantics are described, and some quantitative measures for the rules of the logics are defined. Finally, in Section 7, we discuss the main contribution of this paper and indicate the direction of future research.

2. Classical data tables

In data mining problems, data is usually provided in the form of data tables (DT). A formal definition of a data table is given in Pawlak (1991).