Information Sciences 271 (2014) 93–114

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Approximate distribution reducts in inconsistent interval-valued ordered decision tables

Wen Sheng Du, Bao Qing Hu *

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, PR China

ARTICLE INFO

Article history:
Received 13 November 2012
Received in revised form 23 January 2014
Accepted 11 February 2014
Available online 20 February 2014

Keywords:
Inconsistent interval-valued ordered decision tables
Dominance-based rough sets
Attribute reduct
Discernibility matrix approach
Rule acquisition

ABSTRACT

Many methods based on the rough set theory to deal with information systems have been proposed in recent decades. In practice, some information systems are based on dominance relations and may be inconsistent because of various factors. Moreover, taking the imprecise evaluations and assignments in the description of objects into account, single-valued information systems have been generalized to interval-valued information systems. In this paper, by introducing a dominance relation to interval-valued ordered information systems, we establish a dominance-based rough set approach, which is mainly based on substitution of the indiscernibility relation by the dominance relation. To extract the minimal decision rules, approximate distribution reducts are proposed in inconsistent interval-valued ordered decision tables. This paper presents a theoretical method based on the discernibility matrix to enumerate all reducts and a practical approach on the basis of significance to find one reduct. And two equivalent definitions of approximate distribution reducts are also introduced. In addition, numerical examples are employed to examine the validity of the approaches proposed in this paper.

1. Introduction

The rough set theory (RST), initiated by Pawlak in the early 1980s [35] (see also [37]), serves as an extension of the classical set theory for the study of intelligent systems characterized by insufficient and incomplete information [36]. The starting point of this theory is an observation that objects having the same description are indiscernible in the view of the available information about them. Within thirty years of development, it has been successfully applied in many fields such as machine learning, intelligent systems, inductive reasoning, pattern recognition, knowledge discovery, decision analysis, expert systems.

The original rough set theory is not able, however, to discover inconsistencies coming from consideration of criteria, that is, attributes with preference-ordered domains (scales), such as product quality, market share, and debt ratio. To address this issue, Greco, Matarazzo and Słowiński proposed an extension of the rough set theory, which is called the dominance-based rough set approach (DRSA) to take into account the ordering properties of criteria [14,15]. This innovation is mainly based on the substitution of the indiscernibility relation by a dominance relation, which permits approximation of ordered sets in multiple criteria sorting problems. In DRSA, where condition attributes are criteria

* Corresponding author. Tel.: +86 027 68752966; fax: +86 027 68752256.
E-mail addresses: wsdu@whu.edu.cn (W.S. Du), bqhu@whu.edu.cn (B.Q. Hu).

http://dx.doi.org/10.1016/j.ins.2014.02.070
0020-0255/

and decision classes are preference-ordered, the knowledge approximated is a collection of upward and downward unions of classes and the granules of knowledge are sets of objects defined using a dominance relation [25–27,45,46,54].

Dubois and Prade presented the concept of rough fuzzy sets and fuzzy rough sets [12] by means of integrating the RST with the fuzzy set theory [11]. Likewise, the DRSA could be combined with fuzzy set theory as well. In [16], fuzzy set extensions of the DRSA are proposed to deal with fuzziness in preference representation. Dominance-based fuzzy rough set [13] based on t-norm, s-norm and implication operations serves as an extension of $(\mathcal{I},\mathcal{T})$-fuzzy rough set [44] established in RST and can be used to handle uncertain information. To handle with preference analysis in intelligent data analysis and machine learning, fuzzy preference based rough set approach [21] was presented by Hu et al. which is mainly based on the construction of the fuzzy preference relations from samples characterized by numerical criteria.

From the logical viewpoint, a decision table may also be seen as a series of decision rules. How to simplify these decision rules is one of the major topics discussed in DRSA. As pointed out by Greco et al. [4,14], the set of decision rules induced from the approximations defined using dominance relations gives, in general, a more synthetic representation of knowledge contained in the decision table than the set of rules induced from classical approximations defined using indiscernibility relations. And they are more understandable and more applicable for the users because of the more general syntax of the rules.

According to Inuiguchi et al. [28], there exist at least four sources of inconstancy in decision tables, which are listed as follows: (1) hesitation in evaluation of decision attribute values, (2) errors in recording, measurement and observation, (3) missing condition attributes related to the evaluation of decision attribute values, (4) the unstable nature of the system represented by the decision table and the like. These inconsistencies cannot be considered as a simple error or noise. They may convey important information that should be taken into account. To acquire brief decision rules from inconsistent decision tables, relative attribute reducts are needed. Skowron and Rauszer introduced the discernibility matrix method which became a popular approach to enumerate all reducts in the RST [47]. Susmaga et al. introduced a discernibility matrix to ordered information systems and addressed the computation of reducts in DRSA [49]. Recently, a general definition of discernibility matrices was suggested by Miao et al. to summarize the common structures of the existed ones [33].

In the RST, this problem has already been faced in the literature. The concept of the approximate distribution reduct (μ-decision reduct) was proposed by Kryszkiewicz to deal with inconsistent systems [30]. However, the practical approach to compute all approximate distribution reducts was proposed by Zhang et al. [59]. Using the approximate distribution consistent set, the derived rules are compatible with the ones from the original system, that is to say, if two decision rules derived respectively from the reduced and the original system are supported by a same object, then their decision parts must be the same. The approximate distribution reduct has also been studied in other types of information systems, for instance, inconsistent incomplete decision tables [42]. On the other hand, β-reduct proposed by Ziarko was studied in the variable precision rough set model by reducing boundary area in decision tables [63]. But the derived decision rules from the β-reduct may be in conflict with the ones from the original system. To overcome this kind of drawback, Mi et al. introduced concepts of β lower distribution reduct and β upper distribution reduct [32]. Some algorithms have also been proposed for extracting decision rules from an inconsistent decision table, for example, algorithm REBCA [40].

Parallel work have also been made for the DRSA along this line of research. Nevertheless, only a limited number of methods using the DRSA to acquire knowledge in inconsistent ordered decision tables have been proposed. The distribution reduct and maximum distribution reduct were proposed by Xu et al. to meet different requirements in inconsistent ordered decision tables [51]. Later, Xu et al. further proposed another two types of reducts, possible and compatible distribution reducts [53]. As mentioned before, some mislabeled samples may lead to inconsistent ordered decision tables. It was reported that dominance-based rough sets are heavily sensitive to these noisy samples. Inuiguchi et al. [28] presented a variable-precision dominance-based rough set approach (VP-DRSA) and studied union-based reducts in VP-DRSA. To design a robust metric of feature quality, Hu et al. introduced a metric function, which is called rank mutual information (RMI) for monotonic classification [22]. On the other hand, by introducing of rank entropy, they designed a decision tree technique (REMT) based on RMI [23]. It has been proved that the proposed technique can produce monotonically consistent decision trees if the given training sets are monotonically consistent.

An interval-valued information system is an important type of information system, and a generalization of single-valued information system. Some problems of decision making in the context of interval-valued information systems have been studied in [6,19,20,24,41,55], most of which are based on the concept of a possibility degree between any two interval numbers. Some similarity/distance measures for interval-valued fuzzy sets have been proposed to measure the similarity/difference of interval-valued fuzzy sets [50,60]. The principle of the maximum (minimum) degree of similarity (difference) between interval-valued fuzzy sets can be used to solve problems of pattern recognition and decision making. Precisely, the more (less) the similarity (difference) between the sample and patterns already known, the more likely the sample should be clarified to the pattern.

The knowledge reduction of interval-valued fuzzy information systems is studied under the circumstances of the condition and decision attributes are nominal, not ordinal [17,48]. As for interval-valued ordered information systems (IvOISs), Qian et al. proposed the DRSA to attribute reducts in IvOISs based on a dominance relation for interval numbers [41]. However, they did not mention the underlying concepts of relative attribute reducts in inconsistent interval-valued ordered decision tables (IvODTs) and only proposed an approach to relative attribute reducts in consistent IvODTs. Therefore, the purpose of this paper is to develop approaches to relative attribute reducts in inconsistent IvODTs to make a completion of their work.

The other parts of this paper are organized as follows. In Section 2, the dominance relations in IvOISs are reviewed and some important properties of the dominance classes are also investigated. In Section 3, dominance-based rough sets are