Knowledge-Based Systems 31 (2012) 140–161

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

# Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining

Hongmei Chen [a,b], Tianrui Li [a,b,*], Da Ruan [c,d]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
[b] Key Lab of Cloud Computing and Intelligent Technology, Sichuan Province, Chengdu 610031, China
[c] Belgian Nuclear Research Centre (SCK•CEN), Mol, Belgium
[d] Ghent University, Gent, Belgium

ARTICLE INFO

Article history:
Received 16 May 2011
Received in revised form 27 January 2012
Accepted 3 March 2012
Available online 9 March 2012

Keywords:
Granular computing
Incomplete Ordered Decision Systems (IODSs)
Knowledge discovery
Extended dominance characteristic relation
Approximations

ABSTRACT

Approximations in rough sets theory are important operators to discover interesting patterns and dependencies in data mining. Both certain and uncertain rules are unraveled from different regions partitioned by approximations. In real-life applications, an information system may evolve with time by different factors such as attributes, objects, and attribute values. How to update approximations efficiently becomes vital in data mining related tasks. Dominance-based rough set approaches deal with the problem of ordinal classification with monotonicity constraints in multi-criteria decision analysis. Data missing frequently appears in the Incomplete Ordered Decision Systems (IODSs). Extended dominance characteristic relation-based rough set approaches process the IODS with two cases of missing data, *i.e.*, "lost value" and "do not care". This paper focuses on dynamically updating approximations of upward and downward unions while attribute values coarsening or refining in the IODS. Under the extended dominance characteristic relation based rough sets, it presents the principles of dynamically updating approximations w.r.t. attribute values' coarsening and refining in the IODS and algorithms for incremental updating approximations of an upward union and downward union of classes. Comparative experiments from datasets of UCI and empirical results show the proposed method is efficient and effective in maintenance of approximations.



## 1. Introduction

Granular computing (GrC) proposed by Zadeh [1,2] and Lin [3] has been widely used in image processing, pattern identification, knowledge discovery, and many other fields [4–7]. Granule is a clump of objects drawn together by indistinguishability, similarity or functionality [2]. Granules and their relationships are employed to find solutions to any desired problems. Different granularity levels are formed by decomposing and composing of granules. Different concept or rule levels are then unraveled. The frameworks, models, methodologies, and techniques of GrC were studied in [8–10]. Yao described basic issues and methods of GrC [11–13]. Pedrycz proposed the knowledge-based fuzzy clustering to form granules and further investigated GrC related work, *e.g.*, a granular-oriented neural network, a granular time series approach [14–18].

Rough sets theory is a mathematical tool to process inconsistent data in information systems [19–22]. It is seemed as an important sub branch of GrC. Skowron and Polkowski studied the rough sets theory based GrC in [23–25]. The equivalence relation between any of the two objects is formed by an equality of attribute values in Traditional Rough Sets (TRSs). The preference order is important to the multi-criteria decision analysis, *e.g.*, credit appraisal, risk evaluation, and feasibility study [26,27]. Greco et al. proposed a Dominance-based Rough Set Approach (DRSA) for decision making [28]. The equivalence relation in TRS is replaced by the dominating (dominated) relation in the DRSA. Furthermore, they proposed fuzzy dominance relation-based rough sets by introducing fuzzy logic into the dominance relation [29,30]. To deal with the inconsistency caused by errors in recording, measurement and observation, Inuiguchi et al. proposed a Variable-Precision Dominance relation-based Rough Set Approach (VP-DRSA) and studied the reduction under VP-DRSA [31]. Hu et al. proposed a method for extracting fuzzy preference relations from samples characterized by numerical criteria [32]. Qian et al. extended the dominance relation to interval information systems and set-valued information systems [33,34]. Kotłowski et al. introduced a probability model to deal with the problem of the ordinal classification and proposed a

* Corresponding author at: School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China. Tel.: +86 28 86466426; fax: +86 28 87600743.

E-mail addresses: hmchen@swjtu.edu.cn (H. Chen), trli@swjtu.edu.cn (T. Li), druan@sckcen.be, da.ruan@ugent.be (D. Ruan).

0950-7051/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2012.03.001

stochastic DRSA [35]. Huang introduced a graded dominance relation to dominance interval-valued fuzzy objective information systems and proposed graded dominance interval-based fuzzy objective information systems [36]. The DRSA has been used in multi-criteria classification and ordinal attribute reduction [37–39]. It has been also successfully applied in a variety of fields in decision-making, including bankrupt risk prediction [40], customer behavior prediction [41], Kansei data analysis in product development [42], and IT business value analysis [43].

Real-life applications face often data missing due to various uncertainties. Yang et al. studied three cases of data missing in Incomplete Interval-valued Information Systems (IIISs), *i.e.*, interval-valued data with known lower and unknown upper limits, interval-valued data with unknown lower and known upper limits, and interval-valued data with both unknown lower and unknown upper limits. They discussed the relative reduction in IIIS [44]. Shao and Zhang applied DRSA to reasoning in Incomplete Ordered Information Systems (IOISs) [45]. Yang et al. proposed a similarity dominance relation and studied the reduction under IOIS [46]. Hu and Liu proposed a generalized extended dominance relation model in which the proportion of common attributes whose values are not lost between two objects must greater than a given threshold [47]. Luo et al. proposed a Limited Extended Dominance Relation (LEDR) considering maximum and minimum values in the dominating relation or dominated relation when comparing to lost value [48]. Grzymala-Busse proposed a characteristic relation to process the two cases of data missing, that is, "do not know" and "do not care" [49]. Chen et al. proposed an Extended Dominance Characteristic Relation (EDCR) to process Incomplete Ordered Decision Systems (IODSs) both with these two cases of data missing and considering the proportion of common attribute whose values are not lost [50].

The information system evolves with time, namely, the attributes of the information system, the objects in the universe, or the attribute values of the objects may change. The attribute values may change due to different reasons, *i.e.*, revise the errors and the variation in the hierarchy level of attributes' value domain [51]. Attributes' values have a hierarchy structure are common in the real-life applications. In [52], Hong et al. proposed methods to mine cross-level rules under fuzzy rough sets model. Feng et al. proposed algorithms to mine decision rules from different levels of abstraction [53]. In [54], Chen et al. proposed algorithms for learning decision tree classifiers from data with hierarchical class labels. The attributes' values have preference order in DRSA, *e.g.*, scores of students, evaluation of products. There are few literatures about updating knowledge when the attributes' values vary in IODS. In DRSA, certain rules may be induced from lower approximation. The lower approximation of upper union means if an object $x_i$ is at least as good as another object $x_j$ on all of the considered criteria so $x_i$ should belong to the decision classes not lower than which $x_j$ belongs to. The lower approximation of downward union means if an object $x_i$ is at most as good as another object $x_j$ on all of the considered criteria so $x_i$ should belong to the decision classes not higher than which $x_j$ belongs to. When attributes' values evolve with time, the partial order between $x_i$ and $x_j$ may vary. Then the granule induced by the dominating (dominated) classes and the approximations may alter accordingly. Generally, the computation of approximations is a necessary step in knowledge representation and reduction based on rough sets. Approximations may further be applied to data mining related tasks. Considering the partial order in attributes' value set, we introduce the concept of Attribute Values Coarsening or Refining (AVCR) and multi-level AVCR in IODS. The principle for incrementally updating approximations of upward and downward unions of classes is further studied by analyzing the variation of granularity in IODS. The IF–THEN rules can be finally updated since certain/uncertain rules are induced from lower/upper approximations.

Incremental updating is an effective method in dynamic data processing [55–57] since previous data structure and knowledge may be used effectively to maintain knowledge when an information system varies. The efficiency of knowledge discovery can usually be improved because it does not need to recalculate the whole data. Much research has been focused on maintenance of knowledge. In the case of variation of objects, An et al. updated rules firstly in TRS, which requires new objects consistent to original decision table and new decision classes cannot be added to the table [58]. Shusaku and Hiroshi obtained uncertain rules from clinic database incrementally, but they did not mention how to obtain certain rules [59]. Bang and Zeungnam divided new samples into seven types and proposed an algorithm for updating rules, which is relative to conditional classes and decision classes [60]. Tong and An classified new samples into four types, *i.e.*, confirmative, entirely new, entirely contradiction, and partly contradiction. They presented an incremental learning rule method [61]. Liu et al. proposed a parallel algorithm based on an improved discernibility matrix to extract rules [62]. Guo et al. proposed a method to extract rules incrementally based on the search tree, which does not need to construct the discernibility matrix [63]. Zheng and Wang proposed an effective rule extraction method, RRIA, which updates rules incrementally by adding and pruning the rule tree [64]. Blaszczynski and Slowinski presented an incremental updating algorithm to induce rules based on the Apriori algorithm under variable consistency DRSA [65]. Fan et al. studied the different cases according to the effect to the rule set when the objects are added or deleted. They proposed an algorithm for updating rules through updating SI (Strength Index) in the different cases [66]. Skowron et al. proposed the definition of function approximation and gave the methods to induce the rule set through objects known [67]. Liu et al. constructed the accuracy matrix and the coverage matrix and presented the concept of interesting knowledge. They proposed a method to update interesting knowledge through updating the accuracy matrix and the coverage matrix [68]. In the case of variation of attributes, Li et al. updated approximations when multi-attributes are added or deleted simultaneously under rough sets based the characteristic relation by analyzing the change of lower boundary region [69]. Cheng proposed two incremental methods for fast computing the approximations in rough fuzzy sets, *i.e.*, one starts from the boundary set and the other is based on the cut sets of a fuzzy set [70]. In the case of variation of attributes' values, Chen et al. defined AVCR in TRS and proposed algorithms for updating approximations dynamically by analyzing the evolvement of boundary region and granularity [51]. To our best knowledge, an incremental method for updating approximations under IOIS has not yet been discussed so far. In this paper, we discuss methods for updating approximations of upward and downward unions in the EDCR based rough sets w.r.t. AVCR.

The paper is organized as follows. In Section 2, we review basic concepts of IOIS and the dominance relation and introduce the definition of an extended dominance relation. In Section 3, we propose the concepts of AVCR and multi-level AVCR and further investigate the properties, methods and algorithms w.r.t. AVCR. In Section 4, we employ an example to illustrate the proposed methods for incrementally updating approximations. In Section 5, we verify the algorithms by extensive experiments, and then we analyze and discuss the experimental results. We conclude the paper with the future research directions in Section 6.

## 2. Dominance-based rough set approach for IODS

In this section, we introduce basic concepts of decision systems, including IODS, dominate relation and an EDCR in IODS with missing values [28,50,71].