Information Sciences 299 (2015) 221–242

Contents lists available at ScienceDirect

Information Sciences

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

Fast algorithms for computing rough approximations in set-valued decision systems while updating criteria values ☆

Chuan Luo [a,b], Tianrui Li [a,*], Hongmei Chen [a], Lixia Lu [c]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
[b] Department of Computer Science, University of Regina, Regina, Saskatchewan S4S 0A2, Canada
[c] Department of Information Engineering, Shandong Polytechnic, Jinan 250104, China

A R T I C L E I N F O

Article history:
Received 23 September 2013
Received in revised form 25 August 2014
Accepted 11 December 2014
Available online 17 December 2014

Keywords:
Rough set
Set-valued decision systems
Incremental learning
Knowledge discovery

A B S T R A C T

Rough set theory is advocated as a framework for conceptualizing and analyzing various types of data, which is a powerful tool for discovering patterns in a given data set through a pair of concepts, namely, upper and lower approximations. Strategic behaviors need to be reinforced continuously under the dynamic decision environment where data in the decision process can change over time. Incremental learning is an effective technique to deal with dynamic learning tasks since it can make full use of previous knowledge. Set-valued data, in which a set of values are associated with an individual, is common in real-world data sets. Motivated by the needs of knowledge updating due to the dynamic variation of criteria values in the set-valued decision system, in this paper, we present the updating properties for dynamic maintenance of approximations when the criteria values in the set-valued decision system evolve with time. Then, two incremental algorithms for computing rough approximations are proposed corresponding to the addition and removal of criteria values, respectively. Experimental results show our incremental algorithms work successfully on datasets from UCI as well as artificial datasets, and achieve better performance than the traditional non-incremental method.

1. Introduction

Rough set theory is an important mathematical approach to discovery non-trivial, previously unknown, and potentially useful knowledge from databases characterized by insufficient and incomplete information [24,25]. It has recently received much attention on the research areas in both of the real-life applications and the theory itself. Nowadays, rough set theory has been demonstrated to be useful in many research areas such as knowledge discovery [20,43,31], machine learning [2,21], decision analysis and support [44,49,22], expert system [34], and pattern recognition [32].

Multi-Criteria Decision Making (MCDM) is concerned with structuring and solving decision and planning problems involving multiple criteria, which has become increasingly popular in almost the applications related with decision-making [13]. There have already been several attempts at introducing the use of rough set theory to decision-making. However, the original rough set theory with the indiscernibility relation cannot deal with preferential information (preference-ordered attribute domains and decision classes), which is the most sensitive point of MCDM methods. In order to construct a

☆ This is an extended version of the paper presented at the 2013 IFSA World Congress, Edmonton 2013, Canada.

* Corresponding author.
E-mail addresses: luochuan@my.swjtu.edu.cn (C. Luo), trli@swjtu.edu.cn (T. Li), hmchen@swjtu.edu.cn (H. Chen).

http://dx.doi.org/10.1016/j.ins.2014.12.029
0020-0255/

preference model of the decision maker to support MCDM, Greco et al. proposed an extended rough set approach under a dominance relation, called Dominance-based Rough Set Approach (DRSA), which permits approximation of ordered sets in a multi-criteria sorting [10]. Data processing and knowledge acquisition in the ordered information system is one of the most important applications of DRSA. Since Greco et al. established the DRSA in the single-valued ordered information system, many attempts have been made to extend DRSA in various types of ordered information systems. Sai and Yao generalized the notion of information tables to ordered information tables by adding order relations on attribute values, and formulated the solution for mining ordering rules in ordered information tables [42]. Shao and Zhang proposed an extension of the dominance relation in incomplete ordered information systems, which based on the assumption that all the unknown values are lost [33]. Xu et al. proposed possible and compatible distribution reductions in inconsistent ordered information systems, and constructed the dominance matrix, possible and compatible decision distribution matrices to acquire these two forms of distribution reductions [37]. Yang et al. defined the concept of similarity dominance relation, and introduced four new notions of approximate distribution reducts into inconsistent incomplete ordered information systems [41]. Qian et al. established rough set approaches in interval and set-valued ordered information systems, which can simplify knowledge representation and extract much simpler dominance rules [28,26]. Yang et al. further investigated rough set approach in incomplete interval ordered information systems, and proposed six types of relative reducts to generate the optimal decision rules [40]. Xu et al. introduced an intuitionistic fuzzy relation to ordered information systems, and addressed approaches for approximation reduction in intuitionistic fuzzy ordered information systems [36].

In real-world applications, information systems are dynamic in the sense that (1) its records (objects) are continuously being added or deleted, (2) features (attributes) are frequently being added or removed, and (3) contents (attribute values) are continually being updated over time [29,23]. The recently updated data items can be more "interesting" than those data items accumulated some time ago. Tracking these dynamic changes will benefit in human reasoning and improve the ability of decision making. Incremental learning is regarded as the process of using previously acquired learning results to facilitate knowledge maintenance in the changed database instead of processing the whole database from scratch, which leads to considerable reductions in the execution time when maintaining knowledge [1,52,53]. Based on rough set theory, there have been many attempts to research on knowledge updating in dynamic environment. Dey et al. explored the interrelation between discretization and attribute reduction based on rough set theory, and proposed a dynamic discreduction algorithm to merge the two tasks into a single seamless process [8]. Fan et al. proposed an incremental rule-extraction algorithm to deal with the new added data set when the database increased dynamically [9]. Huang et al. proposed a rough set-based rule induction algorithm to incrementally generate concise and complete alternative decision rules when the database increased dynamically [11]. Yang et al. introduced an incremental updating algorithm for computing attribute core based on discernibility matrix [39]. Xu et al. proposed a dynamic attribute reduction algorithm based on 0–1 integer programming while the multiple objects are added into the information system [38]. Wang et al. developed a dimension incremental strategy for attribute reduction based on three representative entropies, e.g., the complementary entropy, the conditional entropy and the conditional complementary entropy [35]. Chen et al. proposed matrix-based incremental algorithms for updating decision rules in case of the attribute values' coarsening and refining [4].

In this paper, we focus on incremental updating approximations based on rough set theory. Due to the dynamic characteristics of decision environment, criteria values in an decision system will evolve over time under the changing requirement of applications. Knowledge granularity will also exhibit a dynamic change accordingly, which leads to the variation of approximations. Incremental algorithms for updating rough set approximations are proposed, respectively, by considering the addition and removal of criteria values in the set-valued decision systems. Experimental results show that the proposed algorithms have better performance than the non-incremental algorithms.

The rest of the paper is organized as follows. Related works are reviewed in Section 2. Section 3 provides basic concepts of rough set theory and set-valued decision system. In Section 4, the updating mechanisms for computing approximations with the variation of criteria values in set-valued decision system are analyzed, and two incremental updating algorithms are proposed with respect to the addition and removal of criteria values, respectively. In Section 5, an example is shown to validate the effectiveness of the proposed algorithm. Performance evaluations are illustrated in Section 6. Finally, the paper ends with conclusions and further research topics in Section 7.

2. Review of related works

Rough approximations are basic operations in rough set theory and are used as the main tools to deal with vague and uncertain data [45]. Based on rough approximations, the universe can be divided into three pair-wise disjoint regions: the positive region, the negative region and the boundary region. These three different regions implies the idea of three-way decisions: immediate acceptance, immediate rejection and deferred decision, which is the main advantage of decision-theoretic rough set theory compared with traditional binary classifiers, such as decision trees, support vector machines, and neural networks [46,47]. Presently, many researchers have concerned the problem of efficient computing approximations, which is the critical issue of rough set-based data analysis.