Information Sciences 372 (2016) 690–708

Contents lists available at ScienceDirect

Information Sciences

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

Efficient updating rough approximations with multi-dimensional variation of ordered data

Shu Wang[a,b], Tianrui Li[a,b,*], Chuan Luo[c], Hamido Fujita[d]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China
[b] Key Laboratory of Cloud Computing and Intelligent Technique, Chengdu Sichuan Province 611756, China
[c] College of Computer Science, Sichuan University, Chengdu 610065, China
[d] Faculty of Software and Information Science, Iwate Prefectural University, Iwate 020-0693, Japan

ARTICLE INFO

Article history:
Received 3 June 2016
Revised 5 August 2016
Accepted 15 August 2016
Available online 16 August 2016

Keywords:
Rough sets
Dominance-based rough sets
Incremental learning
Knowledge discovery

ABSTRACT

Ordered data widely exists in practical problems. Dominance-based Rough Set Approach (DRSA) is an effective mathematical tool to obtain approximations of concepts and discovery knowledge from ordered data. In this paper, we focus on the dynamic DRSA for the multi-dimensional variation of an ordered information system, and propose a novel incremental simplified algorithm which can efficiently update approximations of DRSA when objects and attributes increase simultaneously. Most of existing algorithms can efficiently deal with the single-dimensional variation of an information system. However, multi-dimensional variations often occur in real dynamic data. That is, the object set, the attribute set or attribute values of an information system often vary simultaneously. Although we can directly use the definitions of approximations, or integrate some single-dimensional incremental algorithms to cope with multi-dimensional variations, this always results in complex algorithm architectures and a large amount of inefficient computation. In our works, by simplifying traditional definitions of DRSA and employing the incremental learning strategy, we develop an algorithm to neglect unnecessary parameters and avoid much redundant computation. Then, we present two different storage schemes for our proposed algorithm to solve the problem of memory consumption. Finally, a series of experiments are conducted to evaluate its efficiency. Experimental results clearly show that for the two-dimensional variation of objects and attributes in an ordered information system, our proposed algorithm is much faster not only than the non-incremental algorithm based on traditional definitions, but also than the integration of two single-dimensional incremental algorithms.

1. Introduction

Rough Set Theory (RST) is an excellent mathematical tool for processing vague, imprecise or uncertain problems in reality besides fuzzy set, probability theory and evidence theory [37,50]. Compared with other uncertainty theories, RST can solve uncertainties by existing data without any priori knowledge. Thus it provides relatively simple interfaces and architectures for data modeling. Recently, RST has been widely applied to many real-world problems in decision making [5,13,26,27], pattern recognition [8,12,32,35,39,40,42] and machine learning [21,25,36,47].

* Corresponding author. Tel: +8602866367458.
E-mail addresses: swang@swjtu.edu.cn (S. Wang), trli@swjtu.edu.cn (T. Li), cluo@scu.edu.cn (C. Luo), HFujita-799@acm.org (H. Fujita).

http://dx.doi.org/10.1016/j.ins.2016.08.044

The equivalence relation is defined in classical rough sets to obtain knowledge from an information system. However, for many information systems from practical problems, there usually exists the preference-order relation in their attribute domains and decision classes. These ordered information systems are unable to be processed by classical rough sets. Therefore, Greco et al. defined the dominance relation to substitute the equivalence relation, and proposed the Dominance-based Rough Set Approach (DRSA) [9,10], which can effectively process the information with preference-order domains. Greco et al. further found that the strict dominance principle in DRSA was unsuitable to the problems in reality, so they developed a variable-consistency model of DRSA by introducing a consistency level [11]. Kotlowski et al. proposed a probabilistic DRSA model to deal with ordinal classification problems which were monotonously constrained [20]. Blaszczynski et al. found that consistency measures used in the definition of rough approximation lack monotonicity properties, so they proposed a new consistency measure which involves a specific scheme of computation of previous consistency measure within the variable-consistency model of DRSA [2]. Dembczynski et al. reformulated the dominance principle and introduced the novel second-ordered approximations to preserve the properties of approximations, reduce decision tables and induce decision rules [7]. Inuiguchi et al. proposed several variable-precision DRSA to attribute reduction, and showed their properties in comparison with previous DRSA models [16]. For the incomplete decision system, Luo et al. introduced a limited DRSA model to obtain higher accuracies of approximations [34]. Yang et al. successively proposed different DRSA models in incomplete ordered information systems and incomplete interval-valued information systems [45,46]. They also introduced a new approach for knowledge acquisition in dominance-based incomplete decision systems [44]. Huang et al. presented the notion of dominance intuitionistic fuzzy information systems and established a new dominance relation defined on the score and accuracy function of intuitionistic fuzzy value [14]. Zhang et al. integrated the variable-precision DRSA model with the inclusion measure, and proposed a general framework to study the interval-valued information system [48]. Susmaga introduced a definition of constructs which was context-based in DRSA [43]. Li et al. addressed the problem about constructing the probabilistic rough set model and the multi-granulation decision-theoretic rough set model in the ordered information systems [24]. Ko and Fujita proposed the notion of evidential probability to express the dominance between attributes, and modified DRSA into an evidential model [19]. Besides the theoretical research, many practical applications based on DRSA were also conducted, such as instruments detection [17], airline market management [28–30], business value evaluation [38], and auction mechanisms design [41]. Recently DRSA was further applied to auditing risk judgement [15], candidate supplier selection [15], water quality analysis [18], and vehicle management [1,15].

RST describes uncertain information by two important concepts, namely the upper and lower approximations. Computing approximations is the key step for data mining and knowledge discovery based on rough sets. There are three essential factors in an information system, namely, the object set, the attribute set and attribute values. Dynamic data is certain to lead to variations of the three factors. Therefore, approximations need to be updated in time. According to static algorithms based on the definitions of rough sets, updating approximations is very time-consuming, because whenever approximations are updated, all data in this information system needs to be processed. If updating approximations can be accelerated, the efficiency of processing data will also be improved. When the object set, the attribute set or attribute values vary individually, many scholars employed the incremental learning strategy and presented a lot of single-dimensional approaches to reduce the time of updating approximations [4,6,31,33,49]. For the individual variation of the object set or attribute set in DRSA, Li et al. respectively proposed corresponding single-dimensional incremental algorithms [22,23]. The algorithm for the variation of objects employs a set-based approach while the algorithm for the variation of attributes employs a matrix-based approach. These two algorithms can efficiently update approximations when objects or attributes are added into an ordered information system. However in real-life problems, the object set or attribute set may vary simultaneously. For example, in instruments detection, when new instruments arrive, they may simultaneously carry new performance indexes which have not been considered before; In water quality analysis, when more water samples are tested, new microorganisms may simultaneously appear, and then corresponding indexes also need to be considered. For these multi-dimensional variations, we may use the definitions to update approximations directly, or we may integrate two single-dimensional algorithms to successively deal with newly added objects and attributes, and then the approximations can also be updated. However, these approaches always result in complex algorithm architectures and inefficient computation. Therefore, the specialized multi-dimensional algorithms for updating approximations of rough sets are needed.

For multi-dimensional variations of an information system, Chen et al. proposed a dynamic two-dimensional algorithm for updating approximations based on the decision-theoretic rough sets when objects and attributes are added simultaneously [3]. They decomposed an information system into different sub-spaces and incrementally updated the equivalence relation in these sub-spaces. They also discussed the approximations of the decision-theoretic rough sets in different levels of granularity. However, there are not any multi-dimensional algorithms for updating approximations of DRSA. Thus in this paper, we manage to design an efficient algorithm to calculate approximations for the simultaneous variation of objects and attributes in DRSA. Our contributions are four-fold: 1) We simplify traditional definitions of DRSA to neglect unnecessary parameters for the updating. 2) We present several new notions, e.g., the dominance feature flag, the dominance feature vector and the dominance feature matrix in DRSA, to reflect the dominance relations between objects of their old attributes. By these notions, we employ the incremental learning strategy to avoid much redundant computation. 3) We propose an incremental simplified algorithm for updating approximations efficiently when new objects and attributes are simultaneously added into an ordered information system. 4) We present two implementations for our proposed algorithm based on two different storage schemes to solve the problem of memory consumption. We also conduct a series of experiments in a number of data sets to evaluate their efficiencies.