Knowledge-Based Systems 50 (2013) 218–233

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

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

Incremental approaches for updating approximations in set-valued ordered information systems ☆

Chuan Luo [a], Tianrui Li [a,*], Hongmei Chen [a], Dun Liu [b]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
[b] School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China

ARTICLE INFO

Article history:
Received 6 October 2012
Received in revised form 17 June 2013
Accepted 18 June 2013
Available online 28 June 2013

Keywords:
Rough set theory
Incremental learning
Knowledge discovery
Set-valued information system
Approximations

ABSTRACT

Incremental learning is an efficient technique for knowledge discovery in a dynamic database, which enables acquiring additional knowledge from new data without forgetting prior knowledge. Rough set theory has been successfully used in information systems for classification analysis. Set-valued information systems are generalized models of single-valued information systems, which can be classified into two categories: disjunctive and conjunctive. Approximations are fundamental concepts of rough set theory, which need to be updated incrementally while the object set varies over time in the set-valued information systems. In this paper, we analyze the updating mechanisms for computing approximations with the variation of the object set. Two incremental algorithms for updating the approximations in disjunctive/conjunctive set-valued information systems are proposed, respectively. Furthermore, extensive experiments are carried out on several data sets to verify the performance of the proposed algorithms. The results indicate the incremental approaches significantly outperform non-incremental approaches with a dramatic reduction in the computational speed.

1. Introduction

Granular Computing (GrC), a new concept for information processing based on Zadeh's "information granularity", is a term of theories, methodologies, techniques, and tools that makes use of granules in the process of problem solving [1,2]. With the development of artificial intelligence, the study on the theory of GrC has aroused the concern of more and more researchers [3–5]. Up to now, GrC has been successfully applied to many branches of artificial intelligence. The basic notions and principles of GrC have appeared in many related fields, such as concept formation [6], data mining [7] and knowledge discovery [8,9].

Rough Set Theory (RST) is a powerful mathematical tool for dealing with inexact, uncertain or vague information [10]. It is also known as one of three primary models of GrC [11]. In recent years, there has been a rapid growth of interest in RST and its applications. It seems to be of fundamental importance to artificial intelligence and cognitive sciences, especially in the areas of machine learning, decision analysis, expert systems, inductive reasoning and pattern recognition [13–16]. The data acquired for rough set analysis is represented in form of attribute-value tables, consisting of objects (rows) and attributes (columns), called information systems [17]. In real-life applications, data in information systems is generated and collected dynamically, and the knowledge discovered by RST need to be updating accordingly [12]. The incremental technique is an effective method to update knowledge by dealing with the new added-in data set without re-implementing the original data mining algorithm [18,19]. Many studies have been done towards the topic of incremental learning techniques under RST. Considering the problem of discretization of continuous attributes in the dynamic databases, Dey et al. developed a dynamic discreduction method based on RST and notions of Statistics, which merges the two tasks of discretization and reduction of attributes into a single seamless process, so as to reduce the computation time by using samples instead of the whole data to discretize the variables [20]. Considering the problem of dynamic attribute reduction, Hu et al. proposed an incremental positive region reduction algorithm based on elementary set, which can generate a new positive region reduction quickly when a new object is added into the decision information systems [28]. From the view of information theory, Wang et al. proposed an incremental attribute reduction algorithm based on three representative entropies by considering changes of data values, which can generate a feasible reduct in a much shorter time. However, the algorithm is only applicable on the case of the variation of data one by one [21]. Furthermore, Wang et al. developed a dimension incremental strategy

* Corresponding author.
E-mail addresses: luochuan@my.swjtu.edu.cn (C. Luo), trli@swjtu.edu.cn (T. Li), hmchen@swjtu.edu.cn (H. Chen), newton83@163.com (D. Liu).

0950-7051/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.06.013

for attribute reduction based on information entropy for data sets with dynamically increasing attributes [22]. Since the core of a decision table is the start point to many existing algorithms of attribute reduction, Yang et al. introduced an incremental updating algorithm of the computation of a core based on the discernibility matrix, which only inserts a new row and column, or deletes one row and updates corresponding column when updating the discernibility matrix [29]. Considering the problem of dynamic rule induction, Fan et al. proposed an incremental rule-extraction algorithm (REA) based on RST, which updates rule sets by partly modifying original rule sets without re-computing rule sets from the very beginning and the proposal approach is especially useful in a large database, since it does not re-compute the reducts/rules that are not influenced by the incremental data set [23]. Nevertheless, alternative rules which are as preferred as the original desired rules might exist since the maximum of strength index is not unique. The REA may lead to non-complete rules, then an incremental alternative rule extraction algorithm (IAREA) was proposed to exclude the repetitive rules and to avoid the problem of redundant rules [24]. Zheng et al. developed a rough set and rule tree based incremental algorithm for knowledge acquisition, which is not only obviously quicker than that of classic algorithm, but also has a better performance of knowledge learned by the proposed algorithm to a certain degree [25]. Liu et al. defined a new concept of interesting knowledge based on both accuracy and coverage of the generated rules in the information system, and presented an optimization model using the incremental matrix for generating interesting knowledge when the object set varies over time [26,27].

The main goal of RST is to synthesize approximations of concepts from the acquired data, which is a necessary step for expressing and reducing incomplete and uncertain knowledge based on RST [30–32]. The knowledge hidden in information systems can be discovered and expressed in the form of decision rules according to the lower and upper approximations [36–39]. In order to resolve the problem of high computation complexity in computing approximations under the dynamic information systems, many incremental updating algorithms have been proposed. Therefore, extensive efforts have been devoted to efficient algorithms for computing approximations. Li et al. presented an incremental method for updating approximations in an incomplete information system through the characteristic relation when the attribute set varies over time, which can deal with the case of adding and removing some attributes simultaneously in the information system [40]. Since the domain of attributes may change in real-life applications, attributes values may be added to or deleted from the domain, Chen et al. proposed the incremental updating approach of approximations while attributes values coarsening or refining in the complete and incomplete information systems [35]. Zhang et al. discussed the change of approximations in neighborhood decision systems when the object set evolves over time, and proposed two fast incremental algorithms for updating approximations when multiple objects enter into or get out of the neighborhood decision table [33]. Li et al. firstly introduced a kind of dominance matrix to calculate P-dominating sets and P-dominated sets in dominance-based rough sets approach, and proposed the incremental algorithms for updating approximations of an upward union and downward union of decision classes [34]. Instead of considering the incremental updating strategies of rough sets, Cheng proposed two incremental methods for fast computing the rough fuzzy approximations, which are established respectively based on the redefined boundary set and the relation between a fuzzy set and its cut sets [41].

However, to our best knowledge, previous studies on incremental computing approximations mainly concerned in the single-valued information systems, but little attention has been paid to the set-valued information systems. Set-valued information systems are an important type of data tables, and generalized models of single-valued information systems [42]. In many practical decision-making issues, set-valued information systems have very wide applications, which can be used in intelligent decision-making and knowledge discovery from information systems with uncertain information and set-valued information. In such systems, some of the attribute values of an object may be set-valued, which are always used to characterize the incomplete information, *i.e.*, the values of some attributes are unknown or multi-values. On the other hand, we often encounter the scenario where the ordering of properties of the considering attributes plays a crucial role in the analysis of information systems. Considering attributes with preference-ordered domains is an important characteristic of multi-attribute decision making problems in practice. Greco et al. proposed the Dominance-based Rough Sets Approach (DRSA) [44,45]. This innovation is mainly based on the substitution of the indiscernibility relation by a dominance relation. Furthermore, Qian et al. established a rough set approach in Set-valued Ordered Information Systems (SOIS) to take into account the ordering properties of attributes in set-valued information systems, and classified the SOIS into two categories: disjunctive and conjunctive systems [43]. Since the characteristics of the set-valued information systems is different from that of single-valued information systems (such as: some of the attribute values for an object are set-valued), the method for knowledge acquisition in the single-valued information systems cannot be applied directly to the set-valued ones. For this reason, the incremental method for updating approximations in the dynamic set-valued information systems is discussed in this paper. In [46], Zhang et al. proposed an incremental method for computing approximations in set-valued information systems under the tolerance relation, when the attribute set varies with time. In this paper, we focus on updating knowledge under the variation of the object set in SOIS. Firstly, we discuss the principles of incremental updating approximations when the objects in the universe change (increase or decrease) dynamically in the conjunctive/disjunctive SOIS. Then two incremental updating algorithms are proposed based on the principles. Finally, the performances of two incremental algorithms are evaluated on a variety of data sets.

The remainder of the paper is organized as follows. In Section 2, some basic concepts of RST in SOIS are introduced. The principles and some illustrated examples for incremental updating approximations with the variation of the object set are presented in Section 3. In Section 4, we propose the incremental algorithms for computing approximations based on the updating principles. Performance evaluations are illustrated in Section 5. The paper ends with conclusions and further research topics in Section 6.

2. Preliminaries

For convenience, some basic concepts of rough sets and SOIS are reviewed in this section [42,43].

A set-valued information system is an ordered quadruple $S = (U, C \cup \{d\}, V, f)$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty finite set of objects, called the universe. C is a non-empty finite set of condition attributes and d is a decision attribute with $C \cap \{d\} = \emptyset$; $V = V_C \cup V_d$, where V is the domain of all attributes, V_C is the domain of all condition attributes and V_d is the domain of the decision attribute; f is a mapping from $U \times (C \cup \{d\})$ to V such that $f: U \times \{C\} \to 2^{V_c}$ is a set-valued mapping and f: $U \times \{d\} \to V_d$ is a single-valued mapping.

In an information system, if the domain (scale) of a condition attribute is ordered according to a decreasing or increasing preference, then the attribute is a criterion.