Information Sciences 178 (2008) 4105–4113

Contents lists available at ScienceDirect

Information Sciences

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

The algebraic structures of generalized rough set theory

Guilong Liu [a,*], William Zhu [b]

[a] School of Information Science, Beijing Language and Culture University, Beijing 100083, China
[b] College of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

Article history:
Received 12 July 2007
Received in revised form 15 June 2008
Accepted 22 June 2008

Keywords:
Rough sets
Generalized rough sets
Granular computing
Approximation operators
Solitary sets

ABSTRACT

Rough set theory is an important technique for knowledge discovery in databases, and its algebraic structure is part of the foundation of rough set theory. In this paper, we present the structures of the lower and upper approximations based on arbitrary binary relations. Some existing results concerning the interpretation of belief functions in rough set backgrounds are also extended. Based on the concepts of definable sets in rough set theory, two important Boolean subalgebras in the generalized rough sets are investigated. An algorithm to compute atoms for these two Boolean algebras is presented.

1. Introduction

The rough set theory, proposed by Pawlak [9,10] as a method for data mining in 1982, has attracted the interest of researchers and practitioners in various fields of science and technology. This technique has led to many practical applications in various areas such as, but not limited to, medicine, economics, finance, engineering, and even arts and culture [16,17]. Combined with other complementary concepts such as fuzzy sets, statistics, and logical data analysis, rough sets have been exploited in hybrid approaches to improve the performance of data analysis tools.

The basic structure of rough set theory [11–13] is an approximation space consisting of a universe of discourse and an equivalence relation imposed thereon. However, equivalence relations are too restrictive for many applications; for instance, in existing databases the values of attributes could be either symbolic or real-valued. Rough set theory would have difficulty in handling such values. To address this issue, several known generalizations of rough set model have been reported in the literature. For example, rough set model is extended to arbitrary binary relations [7,14,23,24,26,27,29,38] and coverings [35–37]. Some researchers have even extended rough sets to Boolean algebra [5,18], completely distributive lattices [2] and residuated lattices [19].

An important generalization of rough set theory is the generalized rough set based on arbitrary binary relations on a universal set. Numerous papers have been published on rough sets. In comparison, however, relatively few results have been obtained for generalized rough sets based on arbitrary binary relations. In the past, studying the algebraic structure of a mathematical theory has proved itself effective in making the applications in the sciences more efficient. This is the inherent motivation for us to study the algebraic structures of these generalized rough sets. Such research may not only provide more insight into rough set theory, but also hopefully develop methods for applications.

* Corresponding author. Tel.: +86 010 82303656.
E-mail address: liuguilong@blcu.edu.cn (G. Liu).

0020-0255/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2008.06.021

A similar investigation was done by Yao [27], but this investigation focused on binary relation-based rough sets with special properties. It is preferable to establish the algebraic structure for binary relation-based rough sets without any constraint on the binary relations. Our aim in this paper is to explore the algebraic structures of the lower and upper approximations of generalized rough sets for general binary relations. In fact, our results extend those given in [27]. In achieving our aim, we use as core concept the solitary set, which is introduced for the first time in this paper. With this approach, we will determine a clearer algebraic structure for generalized rough sets based on binary relations that will allow researchers to better understand this type of rough set.

Applications of binary relation-based rough sets to practical situations can be found in the literature [23,24,32,33]. Rough set theory has also been used to interpret belief functions [20,21,28]. Based on arbitrary binary relations, this paper studies the structures of generalized rough sets and interprets the associated belief functions. We propose two concepts of definable sets for generalized rough sets corresponding to definable sets in classical rough sets. Their lattice structures are investigated and an algorithm is presented to compute the particular atoms thereof. Kondo [6] also studied the structure of generalized rough sets based on binary relations from a topological point of view.

The paper is organized as follows. Section 2 introduces relevant definitions pertaining to rough sets and presents the concept of the solitary set to describe the structure of the lower and upper approximations in the generalized rough set environment. Section 3 is concerned with the interpretation of belief functions in generalized rough sets based on arbitrary binary relations. Section 4 presents the definitions of two concepts of definable sets in generalized rough sets based on arbitrary binary relations, and also an algorithm to compute atoms for two important Boolean subalgebras related to these two concepts. Finally, Section 5 concludes the paper.

2. Basic concepts and properties

In this section, we consider the fundamental properties of generalized rough sets induced by arbitrary binary relations. Using a concept defined in this section, the solitary set, we establish a clear structure of the lower and upper approximations in generalized rough sets based on arbitrary binary relations.

Let U be a non-empty set of objects called the universe. U can be an infinite set, i.e., we do not restrict the universe to a finite one. Let R be an equivalence relation on U. We use U/R to denote the family of all equivalence classes of R (or classifications of U), and $[x]_R$ to denote an equivalence class of R containing the element $x \in U$. The pair (U, R) is called an approximation space. For any $X \subseteq U$, we can define the lower and upper approximations of X [9,10,27] as

$$\underline{R}X = \{x | [x]_R \subseteq X\} \quad \text{and} \quad \overline{R}X = \{x | [x]_R \cap X \neq \emptyset\},$$

respectively. The pair $(\underline{R}X, \overline{R}X)$ is referred to as the rough set of X. The rough set $(\underline{R}X, \overline{R}X)$ gives rise to a description of X under the present knowledge, i.e., the classification of U.

Numerous research papers [1,4,5,8,30,31,35] have pointed out the need to introduce a more general approach by considering an arbitrary binary relation (or even an arbitrary fuzzy relation in two universes) $R \subseteq U \times U$ in the set U of objects instead of an equivalence relation.

Suppose R is an arbitrary relation on U. Then, the pair (U, R) is called an approximation space. With respect to R, we can define the neighborhood of an element x in U as follows:

$$r(x) = \{y | y \in U, xRy\}.$$

The neighborhood $r(x)$ becomes an equivalence class containing x if R is an equivalence relation. For an arbitrary relation R, by substituting equivalence class $[x]_R$ with neighborhood $r(x)$, we define the operators $\underline{R}$ and $\overline{R}$ [27] from $P(U)$ to itself as

$$\underline{R}X = \{x | r(x) \subseteq X\} \quad \text{and} \quad \overline{R}X = \{x | r(x) \cap X \neq \emptyset\}.$$

$\underline{R}X$ is called a lower approximation of X and $\overline{R}X$ an upper approximation of X. The pair $(\underline{R}X, \overline{R}X)$ is referred to as a generalized rough set. The set $\underline{R}X$ consists of those elements whose neighborhoods are contained in X, and $\overline{R}X$ consists of those elements whose neighborhoods have a non-empty intersection with X. Obviously, if R is an equivalence relation, then $r(x) = [x]_R$. These definitions are equivalent to Pawlak's original definitions. Note that the definitions of the lower and upper approximations are not unique. For example, we can use the left neighborhood $l(x) = \{y \in U | yRx\}$ to define the lower and upper approximations. Moreover, the rough set concept can be defined quite generally by means of the two topological operations, interior and closure [15].

The difference between the upper and lower approximations is called the R-boundary of X and is denoted by $BN_R(X)$, i.e.,

$$BN_R(X) = \overline{R}X - \underline{R}X.$$

Definition 1 (*Solitary element*). Given a binary relation R on U, if $x \in U$ and $r(x) = \emptyset$, we call x a solitary element with respect to R. The set of all solitary elements with respect to R is called the solitary set and is denoted by S, i.e.,

$$S = \{x | x \in U, r(x) = \emptyset\}.$$