International Journal of Approximate Reasoning 59 (2015) 68–80

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Matrix-based set approximations and reductions in covering decision information systems

Anhui Tan [a,b], Jinjin Li [a,*], Yaojin Lin [c], Guoping Lin [a]

[a] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, 363000, Fujian, China
[b] School of Mathematical Sciences, Xiamen University, Xiamen, 361005, Fujian, China
[c] School of Computer Science, Minnan Normal University, Zhangzhou, 363000, Fujian, China

ARTICLE INFO

Article history:
Received 21 June 2014
Received in revised form 12 October 2014
Accepted 12 January 2015
Available online 7 February 2015

Keywords:
Rough set
Covering decision information system
Matrix
Set approximation
Reduct
Minimal description

ABSTRACT

In this paper, we propose matrix-based methods for computing set approximations and reducts of a covering decision information system. First, some matrices and matrix operations are introduced to compute the set approximations, and further to compute the positive region of a covering decision system. Second, the notions of minimal and maximal descriptions in a covering decision system are proposed which can be easily obtained by the matrix-based methods. Then the minimal and maximal descriptions are employed to construct a new discernibility matrix. We claim that by using the minimal and maximal descriptions, we can dramatically reduce the total number of discernibility sets that need to be computed in the new discernibility matrix, thus dramatically reducing the computational time for finding all reducts and one optimal reduct of a covering decision system. In the end, several numerical experiments are conducted to examine the efficiency of the proposed methods.

1. Introduction

Rough set theory, proposed by Pawlak [16], provides a systematic approach for classification of data. It has been widely applied in machine learning and patten recognition [2,5,11,17,18,22,28,31]. The main goal of rough set theory is to describe and conceptualize various types of imprecise and uncertain data. There are mainly two factors that affect the description capability of rough set theory: set approximation and knowledge reduction. First, set approximation is fundamentally important in rough set theory. Given a subset of data, two definable sets are induced to approximate the subset, called lower and upper approximations, respectively. Second, under the condition of keeping set approximation unchanged, knowledge reduction is to remove redundant knowledge from the whole data so as to make a decision quickly while preserving or even improving the decision making ability.

Classical rough set theory can be employed to deal with data which generates equivalence classes. However, in many real-word situations, there are attributes of multiple different types in information systems, e.g., missing ones, numerical ones, set-valued ones and interval-valued ones. Classical rough set theory has a theoretical limitation on analyzing these data. To overcome this limitation, many methods have been introduced to generalize the classical rough set theory. One method is to relax equivalence relation to other binary relations, such as similarity relation, dominance relation and tolerance relation [1,7,8,13,20,20,21,29]. The other method is to extend partitions to coverings [3,6,10,14,15,19,23–25,30,33].

* Corresponding author.
E-mail addresses: shujujiegouwang@126.com (A. Tan), jinjinli@mnnu.edu.cn (J. Li), yjlin@mail.hfut.edu.cn (Y. Lin), guoplin@163.com (G. Lin).

http://dx.doi.org/10.1016/j.ijar.2015.01.006
0888-613X/

In fact, there are close relationships between coverings and binary relations. Zhang and Luo [36] pointed out that several covering-based rough sets and relation-based rough sets can be equally transformed to each other. Wang et al. [26] noticed that a kind of covering information system and dominance relation-based information system are isomorphic under certain conditions. Therefore, the covering-based and relation-based rough sets can be studied within a unified framework in many cases. As a generalization of classical rough sets, covering-based rough sets have been attracting more and more attention, especially on the construction of set approximations [9,12,14,19,23,24,32]. Yao and Yao [32] unified and classified various kinds of set approximations in covering-based rough set theory. Another important topic on covering-based rough sets is knowledge reduction. Zhu and Wang [37,38] proposed a reduction method to reduce redundant elements in a covering while preserving set approximation. But this reduction is no longer in line with the original purpose of knowledge reduction in classical rough set theory [4]. Compared with Zhu's reduction on a single covering, Chen and Wang and coworkers [4,27] first defined a covering decision information system and presented a pioneering study on the reduction of this information system. This kind of reduction is to remove redundant coverings from a family of coverings, and it can be seen as a generalization of the reduction in classical rough set theory. For solving this problem, Chen et al. [4] originally proposed a discernibility matrix for computing the reduct of a covering decision information system. But the discernibility matrix is very time-consuming due to its complex computation. To further improvement, Wang et al. [27] constructed a new discernibility matrix, which greatly reduced the computational complexity of the original discernibility matrix. Note that in a covering decision information system the data cannot be simply handled by classical rough set methods, and that because different subsets of data usually overlap, it is necessary to provide more rational and efficient methods for this issue.

The main objective of this paper is to introduce matrix-based methods for computing set approximations and finding all reducts and one suboptimal reduct of a covering decision information system. In previous literatures, Zhang et al. [34,35] used some matrices to compute the set approximations of composite rough sets. But due to the complexity of overlapping data, these matrices cannot work in a covering decision information system. Thus, in this paper, we define several new matrices and matrix operations by which one can efficiently compute set approximations. On the other hand, the notions of minimal and maximal descriptions in a covering decision information system are proposed which can be easily obtained by the proposed matrix-based methods. We claim that by using the minimal and maximal descriptions, we can dramatically reduce the number of discernibility sets that need to be computed in a covering decision system, thus can further decrease the computational complexity for finding all reducts and one suboptimal reduct to a large degree.

The rest of this paper is organized as follows. In Section 2, we review some basic concepts related to a covering decision information system. In Section 3, we introduce some matrices and matrix operations, and use them to compute the approximations of a set, particularly to compute the position region of a covering decision system. In Section 4, we introduce the minimal and maximal descriptions in a covering decision information system which can be easily obtained by the proposed matrix operations. In terms of the minimal and maximal descriptions, a new discernibility matrix of a covering decision information system is constructed, in which the total number of discernibility sets that need to be computed is dramatically reduced. In Section 5, the proposed methods are evaluated by several numerical experiments.

2. Background

In this section, we review basic concepts related to covering-based rough sets and covering decision information systems. More details can be found in [4,27,32,33,37].

Definition 1. Let U be the universe and C be a family of subsets of U. If none subsets in C is empty and $\cup C = U$, then C is called a covering of U; the ordered pair (U, C) is called a covering approximation space.

One can see that a partition of the universe is certainly a covering of the universe.

Each object of the universe is associated with one subset of objects called the neighborhood of the object. The family of all neighborhoods actually generates a granular space which explicitly reflects the relationship among objects.

Definition 2 *(Neighborhood).* Let C be a covering of U. For $x \in U$, denote $C_x = \cap\{K \in C | x \in K\}$ as the neighborhood of x with respect to C. Then $Cov(C) = \{C_x | x \in U\}$ is also a covering of U and we call it the induced covering of C.

Definition 3 *(Neighborhood).* Let Δ be a family of coverings of U. For each $x \in U$, denote $\Delta_x = \cap\{C_x | C \in \Delta\}$ as the neighborhood of x with respect to Δ. Then $Cov(\Delta) = \{\Delta_x | x \in U\}$ is also a covering of U and we call it the induced covering of Δ.

Clearly, Δ_x is the intersection of all the elements including x in the coverings of Δ. If each covering of Δ is a partition, then $Cov(\Delta)$ is also a partition and Δ_x is the equivalence class including x.

Based on the neighborhoods, Chen and Wang and coworkers [4,27] defined the approximation operators as follows.

Definition 4. (See [4,27].) Let Δ be a family of coverings of U. A pair of approximation operators $(\underline{\Delta}, \overline{\Delta})$ is defined as: $X \subseteq U$,