Information Sciences 254 (2014) 181–196

Contents lists available at ScienceDirect

Information Sciences

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

# A novel method for attribute reduction of covering decision systems

Changzhong Wang [a,*], Qiang He [b], Degang Chen [c], Qinghua Hu [d]

[a] Department of Mathematics, Bohai University, Jinzhou 121000, PR China
[b] Department of Mathematics and Computer Science, Hebei University, Baoding 071002, PR China
[c] Department of Mathematics and Physics, North China Electric Power University, Beijing 102206, PR China
[d] School of Computer Science and Technology, Tianjin University, Tianjin 300072, PR China

ARTICLE INFO

*Article history:*
Received 17 April 2011
Received in revised form 6 August 2013
Accepted 24 August 2013
Available online 2 September 2013

*Keywords:*
Covering rough set
Attribute reduction
Discernibility matrix
Consistent covering decision system
Inconsistent covering decision system

ABSTRACT

Attribute reduction has become an important step in pattern recognition and machine learning tasks. Covering rough sets, as a generalization of classical rough sets, have attracted wide attention in both theory and application. This paper provides a novel method for attribute reduction based on covering rough sets. We review the concepts of consistent and inconsistent covering decision systems and their reducts and we develop a judgment theorem and a discernibility matrix for each type of covering decision system. Furthermore, we present some basic structural properties of attribute reduction with covering rough sets. Based on a discernibility matrix, we develop a heuristic algorithm to find a subset of attributes that approximate a minimal reduct. Finally, the experimental results for UCI data sets show that the proposed reduction approach is an effective technique for addressing numerical and categorical data and is more efficient than the method presented in the paper [D.G. Chen, C.Z. Wang, Q.H. Hu, A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets, Information Sciences 177(17) (2007) 3500–3518].



## 1. Introduction

Attribute reduction plays an important role in pattern recognition and machine learning. Classical rough set theory, as proposed by Pawlak [31], has been used as a mathematical tool to conceptualize and analyze various types of imprecise and uncertain data. It can be a useful tool for studying attribute reduction in information systems. The main goal of attribute reduction is to remove redundant information in a data set, so that correct decisions can quickly be made while preserving or even improving classification ability. Rough set theory has become a popular mathematical framework for feature selection, rule extraction, data mining and knowledge discovery [5–23,26–38,44–46,51–53].

Originally, rough set theory was based on equivalence relations, which can partition the objects of a universe into mutually exclusive equivalence classes. Objects in the same equivalence class are indiscernible. Equivalence classes are also called elementary information granules; an arbitrary subset of the universe can be approximated by elementary information granules. Rough set-based data analysis starts from a data table, called an information system, which contains objects that are described by a finite set of categorical attributes [17–22,28–38,44,51,58–60].

However, in many real-world situations, the databases are not suitable for being handled by classical rough sets [1–7,10–27,39–43,45–58,61–68]. For example, some objects in a database have multiple attribute values. If we consider all of the

* Corresponding author. Tel.: +86 15041656439.
*E-mail address:* changzhongwang@126.com (C. Wang).


http://dx.doi.org/10.1016/j.ins.2013.08.057

attribute values for a multi-valued attribute, we create a covering of the universe instead of a partition. Some illustrative examples of this type of database are given by Chen and Wang [5,45]. In addition, classical rough sets work only for categorical data and cannot be directly applied to reducing numerical attributes. Numerical data need to be discretized before attribute reduction, and this causes information loss [14–16]. For this reason, similarity relation rough sets [43], dominance rough sets [12,13,41], and neighborhood rough sets [14,15,55] were developed. These models induce a covering of a universe instead of a partition. Zakowski [64] used coverings of a universe to introduce the notion of covering rough sets. Later, many authors conducted studies on the properties of covering approximation operators [1–5,25,40,45,48–50,57,63–68]. Bonikowski et al. [1] mainly studied the structures of the coverings. Most authors studied some properties of the upper and lower approximations of covering rough sets and examined some axioms that are satisfied by Pawlak rough sets [25,48,63–65,67,68]. However, few people employ covering rough sets to conduct research on attribute reduction. Zhu and Wang [66] investigated the problem of reducing covering elements from a covering. Their objective was to remove excessive covering elements in a covering, under the condition that the approximations of an arbitrary subset were kept unchanged. A pioneering study on attribute reduction with covering rough sets was conducted by Chen and Wang [5]. Formal concepts of attribute reduction based on covering rough sets were introduced, and some approaches to computing all reducts using a discernibility matrix were developed. However, these approaches are very complex and time-consuming due to their poor design.

In this paper, we present a novel method for the attribute reduction of decision systems based on covering rough sets. We review some concepts related to attribute reductions with covering rough sets, present some theorems describing attribute reductions, and then construct discernibility matrices in covering decision systems. Compared with the attribute reduction method presented in [5], the computational complexity of the proposed method is greatly reduced. Finally, based on a discernibility matrix, we develop a heuristic algorithm to find a minimal subset of attributes that approximate an optimal reduct. The experimental results show that the proposed reduction method can effectively handle data and is more efficient than the reduction method presented in [5].

This paper is organized as follows. In Section 2, we recall some basic concepts related to traditional and covering rough sets. In Section 3, we develop a new method of attribute reduction in consistent covering decision systems and examine some basic properties of attribute reduction with covering rough sets. In Section 4, we discuss a theory for reducing the conditional attributes of inconsistent covering decision systems. In Section 5, we present some experiments on some public data sets.

## 2. Preliminaries

First, we review some basic concepts related to classical rough sets that can be found in [30–32].

An information system is a pair $(U,A)$, where $U$ is a nonempty set of samples $\{x_1, x_2, \ldots, x_n\}$, called a universe or a sample space and $A$ is a nonempty set of attributes or features. For any subset $B \subseteq A$, we can define an equivalence relation as follows:

$$\mathrm{Ind}(B) = \{(x,y) \in U \times U : a(x) = a(y), \forall a \in B\}$$

Clearly, $\mathrm{Ind}(B) = \cap_{a\in B}\mathrm{Ind}(\{a\})$. We denote the equivalence class of $x$ with respect to $\mathrm{Ind}(B)$ as $[x]_B$. For $X \subseteq U$, the lower and upper approximations of $X$ with respect to $\mathrm{Ind}(B)$ are defined as follows:

$$\underline{B}(X) = \{x \in U : [x]_B \subseteq X\}, \quad \overline{B}(X) = \{x \in U : [x]_B \cap X \neq \emptyset\}$$

By $M(U,A)$, we denote a $n \times n$ matrix $(c_{ij})$, called the discernibility matrix of $(U,A)$, such that $c_{ij} = \{a \in A\colon a(x_i) \neq a(x_j)\}$ for $i, j = 1, 2, \ldots, n$. A discernibility function $f(U,A)$ of an information system $(U,A)$ is a Boolean function of Boolean variables $\overline{a_1}, \ldots, \overline{a_m}$ corresponding to the attributes $a_1, \ldots, a_m$ and defined as follows:

$$f(A)(\overline{a_1}, \ldots, \overline{a_m}) = \wedge\{\vee(c_{ij}) : 1 \leqslant j < i \leqslant n\}$$

where $\vee(c_{ij})$ is the disjunction operation on $c_{ij}$.

A decision system is a pair $(U,A,D)$, where $D$ is called a decision attribute and the elements in $A$ are called conditional attributes. An attribute $a \in B(\subseteq A)$ is called relatively dispensable in $B$ if $\mathrm{Pos}_B(D) = \mathrm{Pos}_{B-\{a\}}(D)$; otherwise, it is said to be relatively indispensable in $B$, where $\mathrm{Pos}_B(D) = \cup_{X\in U/D}\underline{B}X$. If each attribute in $B$ is relatively indispensable in $B$, $B$ is said to be relatively independent in $(U,A,D)$. A subset $B \subseteq A$ is called a relative reduct in $(U,A,D)$ if $B$ is relatively independent in $(U,A,D)$ and $\mathrm{Pos}_B(D) = \mathrm{Pos}_A(D)$. The set of all relatively indispensable attributes in $A$ is called the relative core of $(U,A,D)$.

Classical rough sets can only be used to address discrete data. Rough sets have been extended to address complex data with discrete and numerical features. Covering rough set theory is a natural extension of classical rough sets and can effectively handle continuous data. Next, we recall some basic notions related to covering rough sets that can be found in [5,45].

**Definition 2.1.** Let $U$ be a universe of discourse and $C$ be a family of subsets of $U$. Then, $C$ is called a covering of $U$ if no subset in $C$ is empty and $\cup C = U$.