Information Sciences 263 (2014) 186–197

Contents lists available at ScienceDirect

Information Sciences

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

Characteristic matrix of covering and its application to Boolean matrix decomposition

Shiping Wang [a], William Zhu [b,*], Qingxin Zhu [a], Fan Min [b]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] Lab of Granular Computing, Minnan Normal University, Zhangzhou 363000, China

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form 22 August 2013
Accepted 16 October 2013
Available online 24 October 2013

Keywords:
Covering
Rough set
Boolean matrix decomposition
Characteristic matrix
Approximation operator

ABSTRACT

Covering-based rough sets provide an efficient means of dealing with covering data, which occur widely in practical applications. Boolean matrix decomposition has frequently been applied to data mining and machine learning. In this paper, three types of existing covering approximation operators are represented by Boolean matrices, and then used in Boolean matrix decomposition. First, we define two characteristic matrices of a covering. Through these Boolean characteristic matrices, three types of existing covering approximation operator are concisely and equivalently represented. Second, these operator representations are applied to Boolean matrix decomposition, which has a close relationship with nonnegative matrix factorization, a popular and efficient technique for machine learning. We provide a sufficient and necessary condition for a square Boolean matrix to decompose into the Boolean product of another matrix and its transpose. We then develop an algorithm for this Boolean matrix decomposition. Finally, these three covering approximation operators are axiomatized using Boolean matrices. This work presents an interesting viewpoint from which to investigate covering-based rough set theory and its applications.

1. Introduction

As a widely used form of data representation, coverings commonly appear in incomplete information/decision systems based on symbolic data [3,4,31], numeric and fuzzy data [14,15,29]. Covering-based rough set theory [30,38] is an efficient tool for dealing with covering data. In recent years, it has attracted much research interest, with a series of significant problems proposed. For example, diverse approximation models have been constructed [5,33,37,43], covering reduction problems have been defined [7,28,34,42], generalizations have been proposed [6,9,18,35,40], and other theories have been combined [11,13,26,32,36].

Specifically, the axiomatization of covering approximation operators has become a popular issue. For example, Zhu and Wang [42,43] proposed a reducible element to axiomatize the covering lower approximation operator. Following Zhu and Wang's work, Zhang et al. [39] axiomatized three pairs of covering approximation operators. Liu and Sai [24] constructed an axiom for a pair of covering approximation operators from the viewpoint of operator theory. Unfortunately, as an efficient tool for computational models, matrices are seldom used in covering-based rough sets. However, an increasing number of achievements have been made in terms of representing and axiomatizing classical and fuzzy rough sets using matrices [22,23,25]. Naturally, this motivates us to represent and axiomatize covering-based rough sets using Boolean matrices.

* Corresponding author. Tel.: +86 189 6526 1496.
E-mail address: williamfengzhu@gmail.com (W. Zhu).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.10.018

Matrix decomposition techniques have attracted increasing attention. These popular techniques include singular value decomposition [8], principle component analysis [16], vector quantization [12], and nonnegative matrix factorization [17]. As a special case of matrix factorization, Boolean matrix decomposition has not only important practical meaning but also profound theoretical significance. In terms of its application, it has been widely used in data mining [2], role engineering [27], and machine learning [10], amongst others. Boolean matrix decomposition problems, such as deterministic column-based matrix decomposition and Boolean matrix factorization, have attracted much theoretical research interest [1,19,20]. The Boolean matrix representation of covering approximation operators also motivates us to design an algorithm for Boolean matrix factorization from the viewpoint of rough sets.

In this paper, we represent three pairs of covering approximation operators using Boolean matrices, and these expressions are in turn used in Boolean matrix decomposition and rough set axiomatization. First, we define two characteristic matrices of a covering, and use them to concisely represent three pairs of covering approximation operators. Second, through the matrix representation of the covering upper approximation operator, we present a sufficient and necessary condition for a square Boolean matrix to decompose into the Boolean product of another Boolean matrix and its transpose. We then design an algorithm to implement this decomposition. It is interesting to note that Boolean matrix decomposition is a special case of matrix factorization, providing various effective and efficient techniques for machine learning research. Third, using the sufficient and necessary condition of Boolean matrix decomposition, we axiomatize these three types of covering approximation operators. Compared with existing results on rough set axiomatization, this work is more interesting and more concise.

The rest of this paper is arranged as follows. Section 2 reviews some fundamental concepts related to covering-based rough sets. In Section 3, we present two types of characteristic matrices of a covering, and use them to represent three types of covering approximation operators. Section 4 exhibits a sufficient and necessary condition, and designs an algorithm, for Boolean matrix decomposition. In Section 5, we axiomatize these three types of covering approximation operators through Boolean matrix decomposition. Section 6 concludes this paper and discusses some avenues for further research.

2. Basic definitions

This section recalls some fundamental definitions and existing results concerning covering-based rough sets.

Definition 1 (*Covering [42]*). Let U be a finite universe of discourse and $\mathbf{C}$ a family of subsets of U. If none of the subsets in $\mathbf{C}$ are empty and $\bigcup \mathbf{C} = U$, then $\mathbf{C}$ is called a covering of U.

Neighborhoods are important concepts in rough sets, characterizing the maximal and minimal dependence between an object and others.

Definition 2 (*Indiscernible neighborhood and neighborhood [41]*). Let $\mathbf{C}$ be a covering of U and $x \in U$. $I_{\mathbf{C}}(x) = \bigcup\{K \in \mathbf{C} | x \in K\}$ and $N_{\mathbf{C}}(x) = \bigcap\{K \in \mathbf{C} | x \in K\}$ are called the indiscernible neighborhood and neighborhood of x with respect to $\mathbf{C}$, respectively. When there is no confusion, we omit the subscript $\mathbf{C}$.

A neighborhood granule derived from coverings is a basic unit for characterizing data. It leads to neighborhood-based decision systems, where neighborhood-based approximation operators have been used extensively in symbolic and/or numeric attribute reduction [14]. In this paper, we study the following three types of lower and upper approximation operators.

Definition 3 (*Approximation operators [30,41]*). Let $\mathbf{C}$ be a covering of U. For all

$$X \subseteq U,$$
$$SH_{\mathbf{C}}(X) = \bigcup\{K \in \mathbf{C} | K \bigcap X \neq \emptyset\}, SL_{\mathbf{C}}(X) = [SH_{\mathbf{C}}(X^c)]^c,$$
$$IH_{\mathbf{C}}(X) = \{x \in U | N_{\mathbf{C}}(x) \bigcap X \neq \emptyset\}, IL_{\mathbf{C}}(X) = \{x \in U | N_{\mathbf{C}}(x) \subseteq X\},$$
$$XH_{\mathbf{C}}(X) = \bigcup\{N_{\mathbf{C}}(x) | N_{\mathbf{C}}(x) \bigcap X \neq \emptyset\}, XL_{\mathbf{C}}(X) = \bigcup\{N_{\mathbf{C}}(x) | N_{\mathbf{C}}(x) \subseteq X\},$$

are called the second, fifth, and sixth upper and lower approximations of X with respect to $\mathbf{C}$, respectively. When there is no confusion, we omit the subscript $\mathbf{C}$.

In summary, the lower approximation of a target set is a conservative approximation, whereas the upper approximation is a liberal approximation based on a priori knowledge. In practical applications, much knowledge is redundant. Therefore, it is necessary to remove the redundancy and retain the essence. For example, the reducible element has been applied to knowledge redundancy in rule learning [7].

Definition 4 (*Reducible element [42]*). Let $\mathbf{C}$ be a covering of U and $K \in \mathbf{C}$. If K is a union of some sets in $\mathbf{C} - \{K\}$, then K is said to be reducible; otherwise, K is irreducible. The family of all irreducible elements of $\mathbf{C}$ is called the reduct of $\mathbf{C}$, denoted as $Reduct(\mathbf{C})$.

3. Matrix representation of covering approximation operators

In this section, we define the matrix representation of a family of subsets of a set, and then propose two types of characteristic matrices of a covering. Through these matrices, we represent three types of covering approximation operators.