Available online at www.sciencedirect.com
ScienceDirect
August 2017, 24(4): 57–68
www.sciencedirect.com/science/journal/10058885
The Journal of China Universities of Posts and Telecommunications
http://jcupt.bupt.edu.cn

Generalized binary discernibility matrix for attribute reduction in incomplete information systems

Ma Fumin[1] (✉), Zhang Tengfei[2]

1. College of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China
2. College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract

To extract and express the knowledge hidden in information systems, discernibility matrix and its extensions were introduced and applied successfully in many real life applications. Binary discernibility matrix, as a representative approach, has many interesting superior properties and has been rapidly developed to find intuitive and easy to understand knowledge. However, at present, the binary discernibility matrix is mainly adopted in the complete information system. It is a challenging topic how to achieve the attribute reduction by using binary discernibility matrix in incomplete information system. A form of generalized binary discernibility matrix is further developed for a number of representative extended rough set models that deal with incomplete information systems. Some useful properties and criteria are introduced for judging the attribute core and attribute relative reduction. Thereafter, a new algorithm is formulated which supports attribute core and attribute relative reduction based on the generalized binary discernibility matrix. This algorithm is not only suitable for consistent information systems but also inconsistent information systems. The feasibility of the proposed methods was demonstrated by worked examples and experimental analysis.

Keywords rough set, generalized binary discernibility matrix, attribute relative reduction, incomplete information system

1 Introduction

In the modern information age, an abundance of data and information exists. Unfortunately, this information explosion has become synonymous with lower efficiency. We are incapable of handling the vast array of information that comes to us; most of the information maybe chaotic, irrelevant and redundant. How to extract and express such implicit knowledge in the form of explicit decision rules hidden in the given information systems has been an active area of research in the past number of decades. While many data analysis approaches were proposed, rough set theory, proposed by Pawlak et al. [1], was shown to be the more promising and has been successfully applied to various areas because of its theoretic simplicity and practicality. By using the concepts of lower and upper approximations in rough set theory, irrelevant and redundant data can be eliminated in the form of knowledge reduction [2].

Attribute relative reduction is some of the most important issue for rough set theory in real-world applications. Various approaches were proposed to compute attribute core and attribute relative reduction from given information systems. The discernibility matrix, developed by Skowron et al. [3], is a highly useful tool for the study of knowledge acquisition and has attracted a great deal of attention from many researchers. Since it has to employ all non-empty elements in the discernibility matrix [4], the discernibility matrix based approaches have always heavy computing load, so its effective improvement is clearly of importance in order to discovery knowledge faster. Binary discernibility matrix was successively developed by Felix et al. [5] to obtain minimal coverings in machine learning based on a binary representation of discernibility between pairs of objects in

Received date: 09-01-2017
Corresponding author: Ma Fumin, E-mail: fmmatj@126.com
DOI: 10.1016/S1005-8885(17)60224-3

a given information system. This beautiful theoretical result makes use of the definition of a binary discernibility matrix to replace sets operations by bitwise operations. Intuitive representation and easy to understand are some of the biggest advantages of binary discernibitlity matrix.

However the binary discernibility matrix is mainly adopted in the complete information system [6]. There are few studies were carried out on the knowledge reduction by using binary discernibility matrix in an incomplete information system. Due to the rampant existence of incomplete information systems in real life, it is therefore necessary to investigate effective approaches to deal with this issue. The potential of the generalized binary discernibility matrix for incomplete information systems was preliminarily presented in Ref. [7]. The work presented in this paper constructs a formulated generalized binary discernibility matrix for knowledge acquisition in incomplete information systems based on extended rough set models. Furthermore, some important properties and judgment theorems are derived, together with some effective and simple algorithms for attribute core computing and attribute reduction in incomplete information system.

The remainder of this paper is structured as follows. In Sect. 2, we summarize some related studies. In Sect. 3, we introduce the major extended rough sets models considered in this paper, including a tolerance relation model, limited tolerance relation model and completed tolerance relation model. In Sect. 4, we contribute a definition of a generalized binary discernibility matrix based upon some extended rough set models in incomplete information system. Some important properties and judgment theorems of the generalized binary discernibility matrix are derived and thereafter an effective knowledge reduction algorithm are developed in Sect. 5. Experiments are conducted to evaluate the efficiency of the proposed algorithms in Sect. 6. Finally, in Sect. 7, we present the conclusions and future work.

2 Related studies

By combining a discernibility matrix with Boolean reasoning techniques, various algorithms were also proposed which compute a reduction of knowledge [2,8]. Hu et al. [9] developed an influential method for computing the attribute core of a decision table by using the discernibility information in the discernibility matrix. Furthermore, Ye et al. [10] found a defect, which may indeed lead to false results for some inconsistent information systems in certain circumstances, and improved the definition of discernibility matrix. To reduce the storage space of the discernibility matrix, Yao et al. [11] derived a reduction construction method based on discernibility matrix simplification in complete decision systems. Yang et al. [12] proposed a heuristic attribute reduction algorithm based on a novel condensing tree (C-Tree), which is a compressed structure of a discernibility matrix, in complete decision systems.

As a powerful approach of knowledge acquisition, binary discernibility matrix has many interesting superior properties and has been rapidly developed to find attribute reduction with several extended rough set models [6,13–19]. Zhi et al. [13] contributed to decreasing the memory space making high efficiency attributes reduction algorithms based on binary discernibility matrix. Ye et al. [14] presented a new binary discernibility matrix to improve the conclusion in the work of Ref. [13]. To overcome the disadvantage of binary discernibility matrices in dealing with variable precision rough set, Chen et al. [15] redefined the traditional binary discernibility matrices, which is called the extensive binary discernibility matrix. Wang et al. [16] studied the upper and lower distribution reduction based on variable precision rough set model and given the upper and lower distribution binary discernibility matrix. Yang et al. [17] further developed binary discernibility matrix in variable precision rough sets to deal with inconsistent information system. However, they still focused on the complete information system.

It is known that Pawlak's classical rough set model can only be used to tackle the problems of complete information systems [20]. However, erroneous, incomplete and uncertain data often present in real-life applications due to data acquisition limitations. Thus, it is inevitable to encounter empty values, which represent inaccessible information in the database at a given instance, in other words, the incomplete information systems with missing values often exist in practical knowledge acquisition [20]. To overcome this limitation, classical rough sets have been extended to several interesting and meaningful general models with respect to differing requirements in the past decades, such as tolerance relation [21–25], valued tolerance relation [26], limited tolerance relation [27], completed tolerance relation [28], neighborhood operators