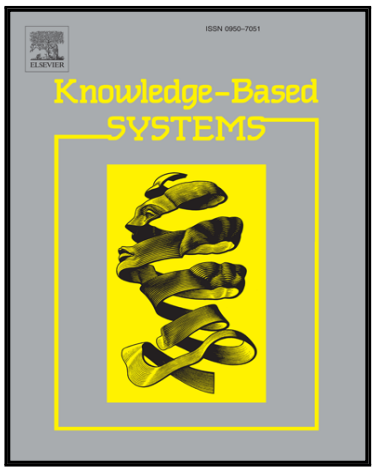

Discernibility matrix based incremental attribute reduction for dynamic data

Wei Wei, Xiaoying Wu, Jiye Liang, Junbiao Cui, Yijun Sun

PII: S0950-7051(17)30502-6
DOI: 10.1016/j.knosys.2017.10.033
Reference: KNOSYS 4095

To appear in: *Knowledge-Based Systems*

Received date: 1 February 2017
Revised date: 27 October 2017
Accepted date: 31 October 2017

Please cite this article as: Wei Wei, Xiaoying Wu, Jiye Liang, Junbiao Cui, Yijun Sun, Discernibility matrix based incremental attribute reduction for dynamic data, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.10.033

Discernibility matrix based incremental attribute reduction for dynamic data

Wei Wei[a], Xiaoying Wu[a], Jiye Liang[a,*], Junbiao Cui[a], Yijun Sun[b,c]

[a]*Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China*
[b]*Department of Microbiology and Immunology, State University of New York at Buffalo, Buffalo, NY14201, USA*
[c]*Department of Computer Science and Engineering, Department of Biostatistics, State University of New York at Buffalo, Buffalo, NY14201, USA*

Abstract

Dynamic data, in which the values of objects vary over time, are ubiquitous in real applications. Although researchers have developed a few incremental attribute reduction algorithms to process dynamic data, the reducts obtained by these algorithms are usually not optimal. To overcome this deficiency, in this paper, we propose a discernibility matrix based incremental attribute reduction algorithm, through which all reducts, including the optimal reduct, of dynamic data can be incrementally acquired. Moreover, to enhance the efficiency of the discernibility matrix based incremental attribute reduction algorithm, another incremental attribute reduction algorithm is developed based on the discernibility matrix of a compact decision table. Theoretical analyses and experimental results indicate that the latter algorithm requires much less time to find reducts than the former, and that the same reducts can be output by both.

Keywords: Attribute reduction; Discernibility matrix; Incremental algorithm; Dynamic data

1. Introduction

Attribute reduction, which is considered an important type of rough set theory based feature selection method [5, 6, 31], aims to select the attributes that retain the discriminatory ability represented by the attribute set of a dataset prior to decision-making [20, 21, 22, 23, 32]. Researchers have proposed numerous algorithms for implementing attribute reductions [1, 5, 26, 33, 34, 37, 38, 40, 41, 46] based on a discernibility matrix, which is a type of representative method [18, 39, 47]. Through a discernibility matrix based attribute reduction, all the reducts can be obtained, which is useful to obtain the minimal reduct and generate a subspace for ensemble learning. It should be noted that most of the algorithms mentioned above are only suitable for static datasets.

However, with the rapid development of information technology, three types of datasets, whose object set, attribute set, or attribute values of objects evolve over time, are ubiquitous in many practical applications [3, 9, 19]. To process these types of datasets, researchers have developed some incremental attribute reduction algorithms over the last two decades. These algorithms are mainly devised based on elementary sets [4], a positive region [28, 30, 42], information entropy and knowledge granularity [7, 14, 36], a discernibility matrix [44, 45], or a dominance matrix [8]. In addition, Xu et al. [43] presented an incremental algorithm based on 0-1 integer programming. The algorithms above were all developed for use with a complete decision table. Zhang et al. [48] provided a matrix representation of the lower and upper approximations in a set-valued information system. Through an analysis of the variations of the relation matrix resulting from the system variance over time, an incremental approach was introduced to update the rough set approximations, through which the updated reducts can be easily obtained. Liu et al. [15, 16, 17] constructed three matrices, based on which some incremental attribute reduction algorithms have been put forward. Chen et al. [2] proposed an equivalent representation of a β-upper (lower) distribution reduct and β-upper (lower) distribution discernibility matrix by means of two Boolean row vectors, and based on these representations, developed a non-incremental algorithm and an incremental algorithm for finding one β-upper (lower) distribution reduct. Nevertheless, most algorithms have aimed at datasets with an object set or attribute set varying over time, and rarely refer to data with attribute values evolving over time. In this paper, we thus focus on attribute reduction for the third type of dataset, i.e., datasets with dynamically varying attribute values, which can be called dynamic datasets [29, 35].

To facilitate the following discussion, we review some possible situations associated with dynamic datasets [36]. One situation is that a dataset has some incorrect values, which need to be replaced to obtain the correct output. Another situation is that data we captured gradually increase in amount over time, although the size of dataset we are interested

*Corresponding author. Tel./Fax: +86 0351 7018176.
Email addresses: weiwei@sxu.edu.cn (Wei Wei), 1652851969@qq.com (Xiaoying Wu), ljy@sxu.edu.cn (Jiye Liang), 945546899@qq.com (Junbiao Cui), yijunsun@buffalo.edu (Yijun Sun)