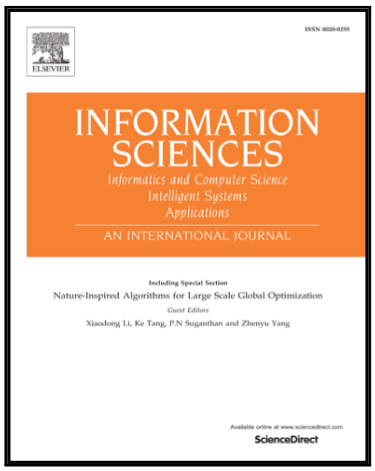

An incremental attribute reduction method for dynamic data mining

Yunge Jing, Tianrui Li, Hamido Fujita, Baoli Wang, Ni Cheng

PII: S0020-0255(18)30519-X
DOI: 10.1016/j.ins.2018.07.001
Reference: INS 13762

To appear in: *Information Sciences*

Received date: 3 February 2018
Revised date: 19 May 2018
Accepted date: 1 July 2018

Please cite this article as: Yunge Jing, Tianrui Li, Hamido Fujita, Baoli Wang, Ni Cheng, An incremental attribute reduction method for dynamic data mining, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.07.001

An incremental attribute reduction method for dynamic data mining

Yunge Jing[a], Tianrui Li[b,*], Hamido Fujita[c], Baoli Wang[a], Ni Cheng[a]

[a]*School of Mathematics and Information Technology, Yuncheng University, Yuncheng 044000, China*
[b]*School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China*
[c]*Faculty of Software and Information Science, Iwate Prefectural University, 020-0693, Iwate, Japan*

Abstract

As an important preprocessing step for data mining, attribute reduction has become a hot research topic in rough set theory. In practice, many real data may vary dynamically with time, therefore, reduct will change dynamically under the variation of objects and attributes in decision systems. The classical attribute reduction methods need to recompute from scratch, which are ineffective to deal with dynamic decision systems. How to implement updating reducts by utilizing previous results is vital for improving the efficiency of attribute reduction approaches. In the paper, we firstly introduce incremental mechanisms of computing reduct when objects and attributes of the decision system change dynamically. Then, incremental methods are developed to update reduct when attributes and objects increase simultaneously. Finally, a series of experiments are conducted to validate the proposed incremental attribute reduction methods. Experimental results show that they are effective to update reduct with change of attributes and objects in the decision systems.

Keywords: Attribute reduction, knowledge granularity, incremental learning, decision system.

1. Introduction

Rough Set Theory (RST) is a useful mathematical method to handle different types of data in decision situations [6, 27]. The model, techniques, frameworks and methodologies of RST have been widely applied in various research areas of decision supporting, classification analysis, knowledge discovery, image processing and intelligent information processing during last decades [1, 28, 30, 32, 33, 38].

Attribute reduction for data sets is a vital research issues in RST[2, 8], which is aimed at deleting the irrelevant or redundant attributes of information systems while keeping its classification ability unchanged. Many attribute reduction approaches have been proposed in recent decades [16, 24, 39]. For example, in a generalized fuzzy-rough set model, Hu et al. developed an effective hybrid attribute reduction method [9]. Qian et al. presented the positive approximation to enhance efficiency of the classical attribute reduction approaches [29]. Miao et al. proposed an approach to obtain reduct of the consistent and inconsistent data sets [25]. But these attribute reduction methods are only suitable for static decision systems. When many attributes and objects change simultaneously in the decision system, the above attribute reduction methods have to recompute the updated decision system from scratch and consume a lot of running time. Hence, they are ineffective to process dynamic data sets.

Nowadays, many decision systems may change over time, i.e., objects, attributes and their values may vary dynamically. How to update reduct is a key issue that can enhance efficiency of data preprocessing. Many incremental updating algorithms have been successfully applied to massive data processing, the interactive applications, and in the case when calculating capacity and memory space are limited. In the case of the variation of attributes, Li et al. designed a dynamic method for updating approximations of rough sets based the characteristic relation in incomplete data sets [17]. Li et al. defined the dominance matrix and designed a dynamic method to update approximations

*Corresponding author.
Email addresses: jyg701022@163.com (Yunge Jing), trli@swjtu.edu.cn (Tianrui Li), HFujita-799@acm.org (Hamido Fujita), pollycomputer@163.com (Baoli Wang), 397670763@qq.com (Ni Cheng)