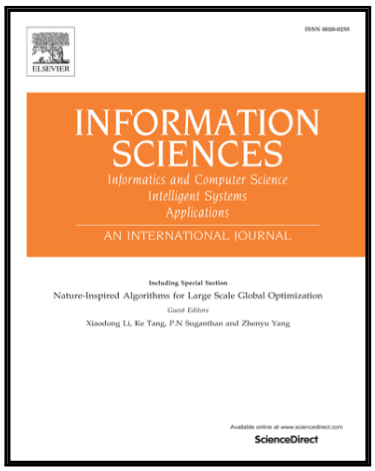

An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view

Yunge Jing, Tianrui Li, Hamido Fujita, Zeng Yu, Bin Wang

PII: S0020-0255(16)30877-5
DOI: 10.1016/j.ins.2017.05.003
Reference: INS 12873

To appear in: *Information Sciences*

Received date: 13 September 2016
Revised date: 7 December 2016
Accepted date: 3 May 2017

Please cite this article as: Yunge Jing, Tianrui Li, Hamido Fujita, Zeng Yu, Bin Wang, An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.05.003

An incremental attribute reduction approach based on knowledge granularity with a multi-granulation view

Yunge Jing[a,b], Tianrui Li[a,*], Hamido Fujita[c], Zeng Yu[a], Bin Wang[a]

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

Abstract

Dynamic updating of attribute reduction is a key factor for the success of rough set theory since many real data vary dynamically with time. Though many incremental methods for updating reduct have been proposed to deal with a dynamically-varying data set and has attracted much attention. However, it is hard to update reduct when the large-scale data vary dynamically. To overcome this deficiency, in this paper, we develop an attribute reduction algorithm with a multi-granulation view to discover reduct of large-scale data sets. Then, incremental mechanisms for knowledge granularity are introduced and two corresponding incremental approaches for updating reduct are developed when many objects are varied in a large-scale decision table with a multi-granulation view. Finally, experiments have been run on six data sets from UCI and the experimental results show that the proposed incremental algorithm with a multi-granulation view can achieve better performance for large-scale data sets.

Keywords: Decision system, incremental learning, knowledge granularity, rough set theory, attribute reduction.

1. Introduction

Rough Set Theory (RST) is a popular mathematical tool for analysis of uncertainty information in decision situations [23]. It has been successfully applied in many research fields such as image processing [6], clustering analysis [8], pattern recognition [36], machine learning [17,30], feature selection [4], decision supporting and data mining [1,40,41]. In RST, attribute reduction is an important concept, which can be known as a kind of specific feature selection. One can select useful features from a given decision system based on RST. Attribute reduction retains the distinguishing power of original decision system for the objects from the universe [11,21].

Attribute reduction is a common technique of preprocessing data and has attracted much attention. In recent years, many attribute reduction methods have been proposed based on RST [5,7,10,27,29,37]. These approaches consume a great deal of computational time and memory space to find a reduct for a large-scale data set. To deal with a large-scale decision system, Slezak presented the model of synthesis of distributed data and methods of evaluating its quality [31]. Huang et al. proposed an incremental approach for updating approximations in distributed information systems under attribute generalization [9]. In addition, Liang et al. proposed an efficient rough feature selection algorithm with a multi-granulation view in order to find a reduct for a large-scale data set [16]. However, the algorithm can only obtain an approximate reduct. To overcome this deficiency, an efficient attribute reduction algorithm with a multi-granulation view is designed in this paper. Firstly, a large-scale decision system is divided into many sub-information systems. A sub-decision system is considered as a small granularity. Then, one can compute an equivalence class on each sub-decision system. Finally, the equivalence class of large-scale decision system is obtained by fusing together these equivalence classes on all sub-decision systems. The total computational time spent on calculating equivalence classes for sub-decision systems is much smaller than that for original large-scale decision system and the memory space

*Corresponding author.
Email addresses: jyg701022@163.com (Yunge Jing), trli@swjtu.edu.cn (Tianrui Li), HFujita-799@acm.org (Hamido Fujita), yuzeng2005@163.com (Zeng Yu), binwang007@gmail.com (Bin Wang)