International Journal of Innovative Computing, Information and Control
Volume **3**, Number **3**, June **2007**

 ISSN 1349-4198
pp. **725–736**

AN EFFICIENT GLOBAL OPTIMIZATION APPROACH FOR ROUGH SET BASED DIMENSIONALITY REDUCTION

Songbo Tan, Xueqi Cheng and Hongbo Xu

Intelligent Software Department
Institute of Computing Technology
Chinese Academy of Sciences
Beijing 100080, P. R. China
tansongbo@software.ict.ac.cn; cxq@ict.ac.cn; hbxu@software.ict.ac.cn

Received January 2006; revised August 2006

Abstract. *The theory of rough set, proposed by Pawlak, provides a formal tool for knowledge discovery from imprecise and incomplete data. Dimensionality reduction (RED) is a crucial problem for rough set based data mining. Unfortunately, it proves to be a NP-hard problem. Many reduction algorithms based on heuristic information have been developed, but they are all based on Boolean space* $\{0,1\}^m$*. In this paper, we introduce the Universal RED problem model, or UniRED, which transforms the discrete attributes reduction problems on Boolean space* $\{0,1\}^m$ *into continuous global optimization problems on real space* E^m*. Based on this transformation, we develop coordinate gradient descent algorithm RED2.1, and coordinate direct descent algorithm RED3.1 for attributes reduction problems. In order to investigate the robustness and efficiency of proposed algorithms, we execute our algorithm RED3.1 on problems from UCI. Meanwhile, the comparison between algorithm RED3.1 and other famous reduct algorithms, such as dynamical reduct algorithm and genetic reduct algorithm, is presented. The experimental results indicate the robustness and efficiency of our algorithms.*

Keywords: Rough set, Dimensionality reduction, Information processing, Knowledge discovery

1. **Introduction.** The rough set theory developed by Pawlak in the early 1980's [1,2] provides a framework for knowledge discovery. Rough set is especially useful for domains where data collected is imprecise and/or incomplete about the domain objects. It provides powerful tools for data analysis and data mining from imprecise and ambiguous data.

Reduct is the most important concept in rough set-based data mining. A reduct is the minimal set of attributes that preserves the indiscernability relation, that is, classification power of the original dataset. The search of a reduct is similar to feature selection problem. All reducts of a dataset can be found by constructing and simplifying the discernibility function [3]. Unfortunately, It has been shown that finding minimal reduct or all reducts are both NP-hard problems.

Therefore, methods to solve these NP-hard problems play an important role in the development of rough set-based data mining. There has been a great interest in designing efficient algorithms to solve these problems. Most algorithms adopt some kind of search algorithm, varying in search direction, search strategy and evaluation measure. These conventional algorithms fall into three categories: the reduction algorithms based on