International Journal of Innovative Computing, Information and Control
Volume **7**, Number **5(A)**, May **2011**
 ISSN 1349-4198
pp. 2167–2186

ATTRIBUTE REDUCTION BY PARTITIONING THE MINIMIZED DISCERNIBILITY FUNCTION

Sirzat Kahramanli, Mehmet Hacibeyoglu and Ahmet Arslan

Department of Computer Engineering
Selcuk University
Konya, Turkey
{ sirzat; hacibeyoglu; ahmetarslan }@selcuk.edu.tr

Received December 2009; revised May 2010

Abstract. *The goal of attribute reduction is to reduce the problem size and search space for learning algorithms. The basic solution of this problem is to generate all possible minimal attributes subsets (MASes) and choose one of them with minimal size. This can be done by constructing a kind of discernibility function (DF) from the dataset and converting it to disjunctive normal form (DNF). Since this conversion is NP-hard, for attribute reduction usually heuristic algorithms are used. But these algorithms generate one or a small number of possible MASes that generally is not sufficient for optimality of dataset processing in such aspects as the simplicity of data representation and description, the speed and classification accuracy of the data mining algorithms and the required amount of memory. In this study, we propose an algorithm that finds all MASes by iteratively partitioning the DF so that the part to be converted to DNF in each of iterations has the space complexity no higher than the square root of the worst-case space complexity of the conversion of the whole DF to DNF. The number of iterations is always fewer than the number of attributes.*
Keywords: Attribute reduction, Feature selection, Discernibility function, Functional partitioning

1. Introduction.

An information system (IS) can be represented as $S = \{U, C, D\}$, where $U = \{O_i\}_{i=1}^{M}$ is a finite set of *objects (instances, records)*, $C = \{A_j\}_{j=1}^{N}$ is a finite set of *condition attributes* and d is a *decision attribute*. Each condition attribute A_j has a *domain of values* $\{A_j(O_i)\}_{i=1}^{M}$, while the decision attribute has a domain of classes $D = \{d(O_i)\}_{i=1}^{M}$, where $A_j(O_i)$ and $d(O_i)$ denote the value of the attribute A_j for the object O_i and the class of this object, respectively. Such an IS is usually considered as a *data table (dataset)* in which the i^{th} row and j^{th} column represent the object $O_i = \{A_j(O_i)\}_{j=1}^{N}$ and the domain $\{A_j(O_i)\}_{i=1}^{M}$, respectively [1]. Generally speaking, the more columns (attributes) there are in a data table, the more memory for its storage and the more time for its processing are needed [2]. Therefore, to reduce the computational difficulties related with processing of large datasets, the *attribute reduction*, known also as *feature selection*, is used.

The goal of attribute reduction is to find a minimal (optimal) subset R of the attribute set C such that R has the same classification power as C. Attribute reduction, allows us to achieve a number of important effects such as: simplification of dataset description, simplification of data collection process, minimization of the needed amount of the data storage, reduction of data classification rules, speeding up data mining algorithms and improvisation of data classification accuracy [2-5]. Therefore, attribute reduction is widely used for preprocessing the datasets used in many fields such as: data mining, decision support systems, knowledge acquisition and discovery, pattern recognition, machine