International Journal of Innovative Computing, Information and Control
Volume **4**, Number **10**, October **2008**

ICIC International ©2008 ISSN 1349-4198
pp. 2749–2757

A NEW DECISION TREE ALGORITHM BASED ON ROUGH SET THEORY

Sang Wook Han and Jae Yearn Kim

Department pf Industrial Engineering
Hanyang University
17 Hangdang-dong, Seongdong-ku, Seoul 133-791, Korea
{softhan; jyk}@hanyang.ac.kr

Received January 2008; revised May 2008

Abstract. *Decision trees are often used in data mining and classification systems because they are easily interpreted, accurate, and fast. One important step in constructing a decision tree is the selection of node attributes so that the tree has a minimum number of branches. While Iterative Dichotomiser 3 (ID3) and C4.5 are the best known of the many decision tree induction algorithms, they do not consider the dependency between entities or the importance of the attributes to the classification. Here, we present two new decision tree classification algorithms, the entity attribute decision tree (EDT) and the reduct attribute decision tree (RDT). These are based on rough set theory and on comparing the values of attributes between entities. We performed experimental comparisons of the new decision tree algorithms with ID3 and C4.5 using the rough set approach to show their advantages of accuracy and rule simplification.*
Keywords: Rough set, Decision tree, Reduct

1. **Introduction.** Classification is one of the most frequent decision making tasks performed by human [1]. Decision trees are commonly used in classification systems [2] because they are easy to interpret, accurate, and fast. The branching decision at each node is determined by a certain splitting criterion that generates a minimal tree, meaning that the tree has a minimum number of branches. The methods currently used for constructing decision trees include ID3, C4.5, CART, and Rough set ID3. Of these, the most popular and efficient are ID3 and its successor C4.5 developed by Quinlan [3,4].

Although the ID3 and C4.5 methods are now well accepted and both attempt to select appropriate attributes using entropy, irrelevant attributes are not filtered until the examples reach the leaf of the decision tree. Those algorithms also have other drawbacks [5–10]. For example, it tends to select the attributes that have more attribute values as the expanding attribute in the process of building decision tree, even if the selected attribute is not the attribute that contributes most to classification. The algorithm may produce an attribute that is individually superior, but when it is selected and used to construct a complete tree, the attribute may result in anomalous behavior and produce an inferior decision tree with poor classification knowledge.

We propose the use of rough set theory introduced by Pawlak in the early 80's for attribute selection in fields, such as data mining, artificial intelligence, and pattern recognition [11–13].

Many variations have been developed and diverse methods have been proposed for the construction of decision trees that use rough set theories, including the attribute core, reduct, and discernibility matrix methods [14-17].