

HEURISTIC MODE RESEARCH OF DECISION TREE ALGORITHM BASED ON SYNTHESIZING EFFECT

FACHAO LI^{1,2}, FEI GUAN² AND YAN SHI^{3,*}

¹School of Economy and Management

²School of Science

Hebei University of Science and Technology
No. 70, Yuhua East Road, Shijiazhuang 050018, P. R. China
lifachao@tsinghua.org.cn; guanfei_118@126.com

³School of Industrial Engineering

Tokai University

9-1-1, Toroku, Kumamoto 862-8652, Japan

*Corresponding author: yshi@ktmail.tokai-u.jp

Received April 2010; revised August 2010

ABSTRACT. *Decision tree, as a simple classification algorithm, is an effective tool for mining knowledge rules, and it has been successfully applied in many fields. Based on the analysis of the essential characteristics of decision tree algorithm for the selection of the expanded attributes, we put forward leaf criterion which can recognize the extension ability of attributes, data utilization criterion and comprehensive effect criterion. Further, we establish a mathematical model with structural characteristics which can evaluate the extension capabilities of attributes, and also give a selection model of expanded attributes based on quasi-linear function (denoted by QASM, for short). Finally, we compare and analyze the performance of QASM combining with ID3 algorithm through different examples. All the results indicate that QASM can effectively merge the decision consciousness into decision process in a quantitative way, and that the computational complexity is lower than that of ID3 algorithm. Thus, it has wide application and operability.*

Keywords: Decision tree, Rule, ID3 algorithm, Expanded attribute, Extension ability, Quasi-linear function, Comprehensive effect

1. **Introduction.** Recently, data mining is increasingly showing its strong vitality. Data classification is an important feature of data mining, and there are many methods for numerical classification such as decision trees, bayesian networks, genetic algorithms and rough sets [1,2]. Decision tree, as a simple classification algorithm, is an effective tool for mining knowledge rules, and it was raised by Hunt [3] in the concept learning system in the early 1960s. Decision tree, using tree structure to represent the decision sets, is a visual knowledge representation, and also a highly efficient classifier. It is very easy to generate association rules using decision tree, and each internal node indicates a test on an attribute; each branch represents a test output; each leaf node is a class or class distribution; the top-level node is the root node. In recent years, data classification technology has been widely and effectively used in scientific experiments, medical diagnosis, weather forecasting, credit, audit, business forecasts, the case detection and so on [4-7].

In 1986, J. R. Quinlan put forward the well-known decision tree induction algorithm ID3 [8], and its main idea is to select the attribute with the largest information gain value from the current samples as the extended attribute by using the information entropy theory, and then determine the division of the sample set according to the value of condition attributes; that is, the number of sub-sample sets depends on the number of the values