



DISCOVERING ACCURATE AND COMMON CHARACTERISTIC RULES FROM LARGE TABLES

YU-CHIN LIU¹ AND PING-YU HSU²

*¹Department of Information Management
Shih Hsin University
Taipei, Taiwan 106, R.O.C.
ycliu@cc.shu.edu.tw*

*²Department of Business Administration
National Central University
Chun-Li, Taiwan 320, R.O.C.
pyhsu@mgt.ncu.edu.tw*

ABSTRACT—With the wide installation of eBusiness and database software in enterprises, mountains of data are accumulating in the form of relational tables. Discovering valuable information from the sea of data is of interest to researchers and managers worldwide. In this paper, an algorithm is proposed to find characteristics from a large database table. It can be applied to find characteristics of customers in a particular segments or the characteristics of patients, ..., etc. In contrast to traditional data generalization or induction methods, the proposed new method, named Char, does not need a concept tree in advance and can generate a manual set of characteristic rules that are precise enough to describe the main characteristics of the data. The simulation results show that the characteristic rules found by Char are efficient as well as consistent regardless of the number of records and of attributes in the dataset.

Key Words: Characteristics rules, Entropy, Redundancy, Information loss, Data mining