# Mining Association Rules in Text

蔣以仁

Pi-Chin Fan;I-Jen Chiang;Ya Wen Tan;Te Chang Huang

## Abstract

In this paper, we propose a new algorithm named Multipass with Inverted Hashing and Pruning (MIHP) for mining association rules between words in text databases. The characteristics of text databases are quite different from those of retail transaction databases, and existing mining algorithms cannot handle text databases efficiently because of the large number of itemsets (i.e., words) that need to be counted. Two well-known mining algorithms, the Apriori algorithm [1] and the Direct Hashing and Pruning (DHP) algorithm [8], are evaluated in the context of mining text databases, and are compared with the proposed MIHP algorithm. It has been shown that the MIHP algorithm has better performance for large text databases.