URN etd-0407109-183705-360
Author Chang Kuo Chen
Author's Email Address No Public.
Statistics This thesis had been viewed 57 times. Download 23 times.
Department
Year 2006
Semester 1
Degree Master
Type of Document Master's Thesis
Language zh-TW.Big5 Chinese
Title Mining Closed Patterns from Long Biological Datasets by Multithread in Multicore CPU
Date of Defense 0000-00-00
Page Count 64

Keyword
Microarray
Data Mining
Frequent Closed Itemset
FP-Growth
Multithread Processing

Abstract
As the bioinformatics technologies growing, the whole human genes have been decoded. The applications of microarray become more and more popular. Some data mining technologies can help us to discover the co-regulated genes or some specific gene groups. But the datasets of microarry are very different from the transactional datasets which are common used in most data mining fields. Basically, the samples of microarray are much less than the record size of transaction datasets and it is generally between 100 and 1,000 data rows. But the number of gene features is very large, it usually reaches thousands or more. So the performance will be bad if we mine those dataset by traditional frequent closed itemset mining algorithms, such as FP-Growth, especially in small minimum support, the run time will increase dramatically. This thesis presents an approach which invokes two or more threads for mining simultaneously, so it will help us to increase efficiency of mining those datasets and alleviates the effect of small minimum support. Besides, this approach can work under multithread environment with duo core or multi core CPU computers which are more popular recently, so the algorithms could work independently to achieve higher performance.
Advisory Committee Hunag, Jen-peng - advisor

Files etd-0407109-183705-360-001.pdf

indicate access worldwide
Date of Submission 2007-08-15