Name of PhD thesis: IMPROVING THE EFFICIENCY OF MINING FREQUENT ITEMSETS BY THE APPROACH OF CLOSED SETS

Speciality: Computer Science   Code: 62 48 01 01

Name of PhD student: Thanh-Trung Nguyen   Course: 3

Supervisor: Associate Professor Nguyễn Phi Khứ

                  Associate Professor Nguyễn Đình Thuân

University of Information Technology – VNU-HCM

Contribution of the PhD thesis:

In Chapter 2, an intermediate structure is called the constructive set P consisting of constructive patterns used to deduce closed sets, frequent sets. The algorithm constructing the constructive set on the direction of increasing transactions is ConPatSet, and the function IncPatSet is to update the constructive set P. When descending or reducing transactions from a transaction set, the function DesPatSet is used to update P.

With a database of n transactions, m items, the complexity of the algorithm ConPatSet and IncPatSet generating P, is indicated by $o(m \times n \times k^2)$, k = |P|.

Storing transaction data in bit-chains has significant advantages. First of all, it reduces the amount of internal and external memory. Then, calculations on bit data are much faster than other types of data. The speed of performing read-write operation between the external and internal memory in bit mechanism is better than in the other mechanisms.

Based on the constructive set, the incremental mining algorithms proposed in Chapter 2 belong to the group of direct-update technique. In Chapter 3, batch processing for incremental mining closed item sets is developed. Specifically, the constructive set of the entire database is generated from component constructive sets, corresponding to dividing the database into transaction batches or item batches. Incremental mining algorithms for transaction or item batch processing are also proposed.

Besides, from the batch processing algorithms, the parallelization solution is developed and deployed on Hadoop-Spark to solve the problems with large volume datasets in Chapter 4.

In addition, the constructive set also is applied to solving the problem of reducing set of characteristic properties and mining data streams.