THESIS INFORMATION

Title:	**Developing effective algorithms for mining frequent patterns and frequent closed patterns from quantitative databases**
Major:	Computer science
Major Code:	62-48-01-01
PhD Student:	Bui Danh Huong
Supervisors:	Assoc.Prof. Vo Đinh Bay Assoc.Prof. Nguyen Hoang Tu Anh
Training place:	University of Information Technology, Vietnam National University – Ho Chi Minh City

1. ABSTRACT

The thesis presents an overview and explores the theoretical basis of mining frequent weighted patterns from quantitative databases. On that basis, the thesis proposes effective methods to solve problems and overcome existing challenges in mining frequent weighted patterns such as the result set is often too large, redundancy rule generation problem and real-time or user-oriented pattern mining. The thesis focuses on solving four specific problems, including mining frequent weighted itemsets, mining frequent weighted closed itemsets, mining top-rank-k frequent weighted itemsets and mining frequent weighted itemsets over data streams. The algorithms proposed in the thesis outperform existing algorithms in terms of running time, memory usage and scalability in solving the above problems.

The research results of this thesis have been published in 4 articles in SCIE journals including Expert System with Application (Q1, 2018), Applied Intelligence (Q2, 2021), Knowledge-Based Systems (Q1, 2020), IEEE Access (Q1, 2021) and in 3 articles in national and international conferences (SMC-2016, FAIR-2016 and @-2017).

2. MAIN CONTRIBUTIONS

The main contributions of the thesis are as follows:

(i) Proposing the WN-tree and WN-list structures, thereby coming up with the NFWI algorithm to efficiently mine frequent weighted itemsets. The WN-list structure has some advantages such as the WN-list intersection has linear complexity, the WN-list has the ability to self-reduce when performing the intersection, and the weighted support of an itemset can be calculated based on the WN-list of that itemset.

(ii) Proposing the pruning strategy based on the WN-list ancestor relation, thereby giving the NFWCI algorithm to efficiently mine frequent weighted closed itemsets based on the WN-list structure.

(iii) Proposing the TFWIN$^+$ algorithm to efficiently mine top-rank-k frequent weighted itemsets based on the WN-list structure and strategies of increasing the threshold and early pruning.

(iv) Proposing SWN-tree structure, an improved version of WN-tree structure, to store and maintain information of data windows when sliding over data stream. From that, the FWPODS algorithm is proposed to efficiently mine frequent weighted itemsets over data streams based on the sliding window model.

3. FURTURE WORKS

In the future, we will focus on solving some problems of pattern mining on quantitative databases such as mining maximal frequent weighted itemsets, mining frequent weighted closed itemsets on incremental databases, mining frequent weighted itemsets on uncertain databases, and implementing frequent weighted itemsets mining on multicore and distributed systems. We also work on implementing applications using the frequent weighted pattern mining platform such as graph mining, social network mining, text data mining, and IoT data mining.

SUPERVISORS **PhD STUDENT**

Assoc. Prof. Vo Đinh Bay Assoc. Prof. Nguyen Hoang Tu Anh Bui Danh Huong