Scalable technique to discover items support from trie data structure

A.Noraziah[a]; Zailani Abdullah[b]; Tutut Herawan[a]; Mustafa Mat Deris[c]
[a]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, Kuantan, Malaysia
[b]Department of Computer Science, Universiti Malaysia Terengganu, Malaysia
[c]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Malaysia

ABSTRACT

One of the popular and compact trie data structure to represent frequent patterns is via frequent pattern tree (FP-Tree). There are two scanning processes involved in the original database before the FP-Tree can be constructed. One of them is to determine the items support (items and their support) that fulfill minimum support threshold by scanning the entire database. However, if the changes are suddenly occurred in the database, this process must be repeated all over again. In this paper, we introduce a technique called Fast Determination of Item Support Technique (F-DIST) to capture the items support from our proposed Disorder Support Trie Itemset (DOSTrieIT) data structure. Experiments through three UCI benchmark datasets show that the computational time to capture the items support using F-DIST from DOSTrieIT is significantly outperformed the classical FP-Tree technique about 3 orders of magnitude, thus verify its scalability.

KEYWORDS:

Frequent Pattern Tree; Trie Data Structure; Fast Technique

REFERENCES

1. Han, J., Pei, H., Yin, Y.: Mining Frequent Patterns without Candidate Generation. In: Proceeding of the 2000 ACM SIGMOD, pp. 1–12 (2000)

2. Zheng, Z., Kohavi, R., Mason, L.: Real World Performance of Association Rule Algorithms. In: Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 401–406. ACM Press (August 2001)

3. Han, J., Pei, J.: Mining Frequent Pattern without Candidate Itemset Generation: A Frequent Pattern Tree Approach. Data Mining and Knowledge Discovery 8, 53–87 (2004)

4. Agrawal, R., Imielinski, T., Swami, A.: Database Mining: A Performance Perspective. IEEE Transactions on Knowledge and Data Engineering 5(6), 914–925 (1993)

5. Liu, G., Lu, H., Lou, W., Xu, Yu, J.X.: Efficient Mining of Frequent Patterns using Ascending Frequency Ordered Prefix-Tree. Data Mining and Knowledge Discovery 9, 249–274 (2004)