Abstract

The running time of existing algorithms in Frequent Pattern Mining (FPM) increases exponentially with increasing average data size. The existing algorithms on high dimensional datasets create large number of frequent patterns of small and mid sizes which are ineffective for decision making and shows deficiency on mining process. To discover large patterns or Colossal Patterns Doubleton Pattern Mining (DPM) is considered as very constructive for analyzing these datasets. In this paper, DPM, An integrated approach for discovering Colossal Pattern from Biological datasets is discussed. DPM effectively discovers a set of Colossal Patterns using vertical top-down column intersection operator. DPM makes use of a data structure called D-struct as combination of a doubleton data matrix and one dimensional array pair set to dynamically discover Colossal Patterns from Biological datasets. D-struct has a diverse feature to facilitate is, it has extremely limited and accurately predictable main memory and runs very quickly in memory based constraints. The algorithm is designed in such a way that it enumerates D-struct matrix iteratively and constructs a phylogenetic tree to discover colossal patterns and takes only one scan over the database. The empirical analysis on DPM shows that, the proposed approach attains a better mining efficiency on various Biological datasets and outperforms Colossal Pattern Miner (CPM) in different settings.
References

Mining (KDD), 2003.

Index Terms

Computer Science Information Sciences

Keywords

Colossal pattern Doubleton Pattern Mining Gene Association Analysis Biological data Colossal Pattern Miner (CPM).