Abstract

Frequent sub graph mining is another active research topic in data mining. A graph is a general model to represent data and has been used in many domains like chemo informatics and bioinformatics. Mining patterns from graph databases is challenging since graph related operations, such as sub graph testing, generally have higher time complexity than the corresponding operations on item sets, sequences, and trees. We investigated new approaches for frequent graph-based pattern mining in graph datasets and found that a novel algorithm called span (graph-based Substructure pattern mining), has been used as the standard for comparing performance of new algorithms. It is based on the pattern growth approach of frequent sub graph mining and hence discovers frequent substructures without candidate generation. span is based on a lexicographic order, and maps each graph to a unique minimum DFS code as its canonical label. Based on this lexicographic order, gSpan adopts the depth-first search strategy to mine frequent connected sub graphs efficiently. Based on the literature survey done, algorithms based on pattern growth approach and DFS strategy are found to be better in performance than algorithms based on Apriori approach and BFS strategy. In this paper we propose a new algorithm based on gSpan, for a special class of graphs characterised by the existence of unique node labels.
References

- X. Yan and J. Han. gSpan: Graph-based substructure pattern mining. Proc. 2nd IEEE Int'l Conf.
- Yuhua Li Quan Lin Gang Zhong Dongsheng Duan Yanan Jin Wei Bi, A Directed Labeled Graph Frequent Pattern Mining Algorithm based on Minimum Code. In the proceedings of Third International Conference on Multimedia and Ubiquitous Engineering 2009.
- Chang Hun You, Lawrence B. Holder and Diane J. Cook: Graph-based Data Mining in Dynamic Networks: Empirical Comparison of Compression-based and Frequency-based Subgraph Mining in International Conference on Data Mining Workshops IEEE, 2004.
- Hsun-Ping Hsieh, Cheng-Te Li, Mining Temporal Subgraph Patterns in Heterogeneous Information Networks: In the proceedings of IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust.
- Bianca Wackersreuther, Peter Wackersreuther, Annahita Oswald: Frequent Subgraph Discovery in Dynamic Networks, ACM 978-1-4503-0214-2 2010.
- M. Kuramochi and G. Karypis. Frequent Subgraph Discovery. In ICDM&apos;01.
- Yong Liu, Jianzhong Li, Hong Gao. JPMiner: Mining Frequent Jump Patterns From Graph Databases. In the proceedings of Sixth International Conference on Fuzzy Systems and Knowledge Discovery 2009.
- Nijssen, S. and Kok, J., Faster association rules for multiple relations. In

Index Terms

Computer Science

Algorithms

Keywords

Parallel programming  frequent subgraph mining  DFS code  Isomorphism