Abstract

In this paper, the problem of finding sequential patterns from graph databases is investigated. Two serious issues dealt in this paper are efficiency and effectiveness of mining algorithm. A huge volume of sequential patterns has been generated out of which most of them are uninteresting. The users have to go through a large number of patterns to find interesting results. In order to improve the efficiency and effectiveness of the mining process, constraints are more essential. Constraint-based mining is used in many fields of data mining such as frequent pattern mining, sequential pattern mining, and subgraph mining. A novel algorithm called CSGP (Constraint-based Sequential Graph Pattern mining) is proposed for mining interesting sequential patterns from graph databases. CSGP algorithm is revised to mine top-k closed patterns and named as TCSGP (Top-k Closed constraint-based Sequential Graph Pattern mining).
5. Feida Zhu, Xifeng Yan, Jiawei Han, and Philip S. Yu, “gPrune: A Constraint Pushing Framework for Graph Pattern Mining”, Advances in Knowledge Discovery and Data Mining, Lecture notes in Computer Science, Springer, Volume 4426, 2007, 388-400.
10. Jian Pei and Jiawei Han, “Can We Push More Constraints into Frequent Pattern Mining?”. In Proceedings of the Sixth ACM SIGKDD international conference on knowledge discovery and data mining, 2000, 350-354.
11. Jian Pei, Jiawei Han, and Laks V.S. Lakshmanan, “Mining Frequent Itemsets with Convertible Constraints”. In proceedings of 11th international conference on Data Engineering, IEEE, April 2001, 433-442.
12. Jian Pei, Jiawei Han, and Wei Wang, “ Mining Sequential Patterns with Constraints in Large Databases”. In Proceedings of CIKM’02 Eleventh International conference on Information and knowledge management, ACM, Newyork, 2002, 18-25.
15. Luc De Raedt, and Albrecht Zimmermann, “Constraint-Based Pattern Set Mining”. In proceedings of the 2007 SIAM International conference on Data Mining, 2007.
17. Marek Wojciechowski and Maciej Zakrzewicz, “Dataset Filtering Techniques in Constraint-Based Frequent Pattern Mining”. In Proceedings of the ESF Exploratory workshop
23. Synthetic graph generated by IBM Quest Synthetic Data Generation Code for Associations and Sequential Patterns. [http://www.7.ust.hk/graphgen/].
25. Wei Wang, Chen Wang, Yongtai Zhu, Baile Shi, Jian Pei, Xifeng Yan, and Jiawei Han, “GraphMiner: A Structural Pattern-Mining System for Large Disk-based Graph Databases and Its Applications”. In proceedings of the 2005 ACM SIGMOD international conference on Management of data, ACM, Newyork, 2005, 89-881.
26. Xifeng Yan, X.Jasmine Zhou, and Jiawei Han, “Mining Closed Relational Graphs with Connectivity Constraints. In proceedings of KDD’05, ACM, Newyork, 2005, 324-333.

Index Terms

Computer Science Applied Mathematics

Keywords

Sequential patterns, Closed patterns, Constraints.