Abstract

Distributed sequential pattern mining is the data mining method to discover sequential patterns from large sequential database on distributed environment. It is used in many wide applications including web mining, customer shopping record, biomedical analysis, scientific research, etc. A large research has been done on sequential pattern mining on various distributed environments like Grid, Hadoop, Cluster, Cloud, etc. Different types of sequential pattern mining can be performed are sequential patterns, maximal sequential patterns, closed sequences, constraint based and time interval based sequential patterns. This paper presents a systematic review on work done for sequential pattern mining and advanced sequential pattern mining on distributed environment. This paper finally presents future research directions related to sequential pattern mining in distributed environment.

References

of the International Conference on Data Engineering (ICDE), Taipei, Taiwan.
- Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan, 2007. "Frequent pattern mining: current status and future directions". Data Mining Knowledge Discovery.
- Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Jianyong Wang, Helen Pinto, Qiming Chen, Umeshwar Dayal, Mei-Chun, 2004. Mining Sequential Patterns by Pattern-Growth : The PrefixSpan Approach. IEEE Transaction on Knowledge and Data Engineering.
- Pei J, Han J, Mao R, 2000. CLOSET: an efficient algorithm for mining frequent closed itemsets. in Proceedings of ACM-SIOMOD international workshop data mining and knowledge discovery (DMKD'00), Dallas.
- Bayardo RJ, 1998. Efficiently mining long patterns from databases. in Proceedings of ACM-SIGMOD international conference on management of data (SIGMOD'98), Seattle, WA.
- Yan, X., Han, J., and Afshar, R. , 2003. CloSpan: Mining closed sequential patterns in large datasets. Third SIAM International Conference on Data Mining (SDM), San Fransisco, CA.


Distributed Sequential Pattern Mining: A Survey and Future Scope


Index Terms

Computer Science Distributed Systems

Keywords

Distributed Sequential Pattern Mining Maximal Patterns Constraint based Patterns

Distributed environment.