Abstract

Web Access Pattern (WAP) tree mining is finding of sequence pattern from web access log. It has gained importance in view of increasing usage of World Wide Web. Access to web pages
generates access log wherefrom meaningful information is extracted. WAP stores web accesses in a prefix tree. In order to mine data, this tree is recursively traversed in bottom up fashion for frequent items that starts with suffix sequences. Repeated construction of sub-trees for finding frequent itemset is necessary in this method. This paper proposes an improved technique termed as WRDSP (WAP Related Dotted Sequence Path) for creation of such graph in which each item needs to be constructed only once. For each attribute, single node only needs be created in proposed approach whereas many nodes may be required for each attribute in conventional WAP approach. To mine frequent pattern from such graph does not require repeated traversal of links already traversed, which is a big saving in memory and time.

References

- Pei, J., Han, J., Mortazavi-Asl, B., and Zhu, H., “Mining access patterns efficiently from web logs”, In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00) Kyoto Japan, 2000.
- Han, J., Pei, H., and Yin, Y, “Mining Frequent Patterns without Candidate Generation”, In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX). ACM Press, New York, NY, USA 2000.
- Lu, Y. and Ezeife C., “Position coded pre-order linked wap-tree for web log sequential

Index Terms

Computer Science  Data Mining

Keywords

Association Rule Mining (ARM)  Frequent Item sets (FIS)
Web Access
Web Access Pattern Relative Dotted Sequence Path (WRDSP)
Web Access Mining (WAM)