Abstract

PL-WAP tree mining algorithm was a great improvement over the WAP tree algorithm which was earlier used for mining in the server web logs. This was basically done to find out frequent sequence patterns. In WAP-tree mining, we used to engage in continuous reconstruction of intermediate WAP trees for each level of recursion to extract frequent patterns. In the WAP technique we could not formulate a method by which we could infer which sequences would be the suffix sequences of a given last event. This was the major drawback of the algorithm. In order to solve this problem, binary codes were introduced to uniquely represent the exact position of each and every node in the WAP-tree. Whereas PLWAP tree is not efficient when the server log data gets frequently updated. Therefore incremental data mining approach is needed to handle this problem. PL4UP proposed later is an incremental mining methodology added to the PLWAP. But the numerous parameters considered, sets generated, insertions and deletions performed at each stage makes it less efficient. and it works efficiently only when the percentage updation in size is less than fifty. In this paper, a novel method is proposed to efficiently handle the updations right from the pre-processing stage and the key factors in the implementation of the method are described. A web usage mining tool was also developed to experimentally validate the method.


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.


Index Terms

Computer Science

Artificial Intelligence

Keywords

Web mining incremental data mining frequent nodes