Abstract

This paper continues the line of research on Web access log analysis. Web access log analysis is to analyze the patterns of web site usage and the features of users' behavior. It is the fact that the normal Log data is very noisy and unclear and it is vital to preprocess the log data for efficient web usage mining process. Preprocessing is the process comprises of three phases which includes data cleaning, user identification and session construction. Session construction is very vital and numerous real world problems can be modeled as traversals on graph and mining from these traversals would provide the requirement for preprocessing phase. On the other hand, the traversals on unweighted graph have been taken into consideration in existing works. This paper oversimplifies this to the case where vertices of graph are given weights to reflect their significance. The proposed method constructs sessions as a Propositional Directed Acyclic Graph (PDAGs) which contains pages with calculated weights. We identify a new property called simple-negation, which is an implicit restriction of all Negation Normal Form (NNFs) and Binary Decision Diagram (BDDs). The removal of this restriction leads to Propositional Directed Acyclic Graphs (PDAG), a more general family of graph-based languages for representing Boolean functions or propositional theories. This will help site administrators to find the interesting pages for users and to redesign their web pages. After weighting each page according to browsing time a PDAGs structure is constructed for each
user session. Existing system in which there is a problem of learning with the Boolean function and the problem can be overcome by the proposed method.

References

- Seong Dae Lee, Hyu Chan Park, “Mining Weighted Frequent Patterns from Path Traversals on Weighted Graph “, IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.4, April 2007.
A Novel Preprocessing Technique for Session Construction using Propositional DAGs

- Theint Theint Aye. 2011. Web Log Cleaning for Mining of Web Usage Patterns IEEE

Index Terms

Computer Science  Information Systems

Keywords

Web Usage Mining  Propositional DAGs  Session Construction  Directed Acyclic Graph (DAG)
Preprocessing
Robots Cleaning