Abstract

Web data mining is an important research topic because it attains a significant amount of interest from both academic and industrial environments. Web sequential pattern mining is an imperative for analyzing the access behavior of web users. Recently, few research works have been designed for mining the web sequential patterns. However, performance of existing techniques was not effectual. In order to overcome such limitation, Hilbert Space clustering based Chronological Backward Search (HSC-CBS) Technique is proposed. HSC-CBS Technique is designed in order to improve the performance of web sequential patterns mining. The HSC-CBS Technique at first used Hilbert space clustering in order to group the similar user’s interest web patterns in web log database which resulting in improved clustering accuracy. The clustering of frequent web patterns in web log database helps for minimizing the space and time complexity of web sequential pattern mining. After clustering, HSC-CBS Technique applied chronological backward search algorithm in order to efficiently mine the web sequential patterns and improving true positive rate of web sequential pattern mining. The HSC-CBS Technique conducts the experimental works on the parameters such as execution
time, space complexity, clustering accuracy, true positive rate of mining and scalability. The
experimental results show that the HSC-CBS Technique is able to improve the true positive rate
of pattern mining and also reduces the execution time as compared to state of the art works.

References

1. Dawei Liu, Saifeng Cai, Xiaohong Guo, “Incremental sequential pattern mining algorithms
   of Web site access in grid structure database”, Neural Computing and Applications, Springer,
   Volume 28, Issue 3, Pages 575–583, March 2017
2. Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng, Clement Yu, “Annotating Search Results
   from Web Databases”, IEEE Transactions On Knowledge And Data Engineering, Volume 25,
   Issue 3, Pages 514 – 527, March 2013
3. C. I. Ezeife, Yi Liu, “Fast incremental mining of web sequential patterns with PLWAP
   tree”, Data Mining and Knowledge Discovery, Springer, Volume 19, Issue 3, Pages 376–416,
   December 2009
   Pattern Mining Based on Session Filter and Transaction Identification”, Journal Of Networks,
   Volume 5, Issue 9, Pages 1018-1024, September 2010
5. Kuo-Wei Hsu, Efficiently and Effectively Mining Time-Constrained Sequential Patterns of
   Smartphone Application Usage”, Hindawi Mobile Information Systems, Volume 2017, Article ID
   3689309, Pages 1-18, 2017
6. Xiuming Yu, Meijing Li, Taewook Kim, Seon-phil Jeong and Keun Ho Ryu, “An Application
   of Improved Gap-BIDE Algorithm for Discovering Access Patterns”, Hindawi Publishing
   Corporation, Applied Computational Intelligence and Soft Computing, Volume 2012, Article ID
   593147, Pages 1-7, 2012
   On PLWAP”, International Journal For Technological Research In Engineering, Volume 4, Issue
   7, Pages 1015-1016, March-2017
   Pattern Using Graph Based Techniques”, International Journal of Scientific & Technology
   Research Volume 3, Issue 7, Pages 76-79, July 2014
    Algorithm Linked WAP-Tree”, Journal of Software, Volume 10, Number 10, Pages 1228-1234,
    October 2015
    Effective Web Page Recommendations”, International Journal of Computer Science and
12. Xiuming Yu, Meijing Li, Kyung Ah Kim, Jimoon Chung and Keun Ho Ryu, “Emerging
    Pattern-Based Clustering of Web Users Utilizing a Simple Page-Linked Graph”, Sustainability,
    Volume 8, Pages 1-18, 2016
13. Meijing Li, Xiuming Yu, Keun Ho Ryu, “MapReduce-based web mining for prediction of
    web-user navigation”, Journal of Information Science, Volume 40, Issue 5, Pages 557-567,
    2014


21. Amazon Commerce reviews set Data Set: https://archive.ics.uci.edu/ml/datasets/Amazon+Commerce+reviews+set

**Index Terms**

Computer Science  
Pattern Recognition

**Keywords**

Chronological backward search, Hilbert Space clustering, mining, web user, web log database, web pattern.