

An Analysis of User Behavior using Closed Set of Agglomerative Approach with GRC Constraints

Amit Kumar Chouksey, Mayank Namdev

Abstract: In the current scenario every Organization need to understand their customers' behavior, preferences and future needs which depend upon past behavior. Web Usage Mining is an active research area in which customers session clustering is done to find out the customer's activities. It investigates the problem of mining frequent pattern and especially focuses on reducing the number of rules using closed pattern technique. It also reduce scans the size of the database using Agglomerative clustering technique using partial database scan. It is perform by Profile based Closed Sequential Pattern Mining with Agglomerative Clustering. It searches the next request page in advance using only partial web data not in whole web data. There is an advantage to no need take input as number of cluster. So it utilized a personalized weighted recommendation system based on user's interest with less execution time.

Keywords: Web Usage Mining, Prefix Span, Gap, Recency, Compactness, Data Stream, Closed Pattern, Data Mining, Personalization, Sequential Pattern Mining, Web Services, Agglomerative Clustering.

I. INTRODUCTION

The popular medium of publishing is the World Wide Web is very rich source of information gathering. It making sense of data is difficult because publication on the web is largely unorganized. Web mining is also knowledge extraction techniques which discover access patterns from the web. It is divided into three parts, a) web usage mining, b) web structure mining and c) web content mining. The commonly used data mining algorithms are Association Rule Mining (ARM), Sequential Pattern Mining, Clustering, and Classification. An ARM technique is used to find out the rules between items found in a transaction database. In the context of web usage mining a transaction is a group of web page accesses with an item being a single page access. The problem of discovering sequential patterns is that of finding inter-transaction patterns such that the presence of a set of items is followed by another item in the time-stamp ordered transaction set. The data mining algorithms are used to generate the association rule between the items, sequential pattern of access of items, and clustering of items. Web Usage Mining (WUM) is the application of data mining techniques to large Web data repositories in order to produce results that can be used in the design tasks and improve response time.

Manuscript published on 30 December 2017.

* Correspondence Author (s)

Amit Kumar Chouksey*, Sagar Institute of Research & Technology (SIRT), Bhopal (Madhya Pradesh), India. E-mail: amit.chouksey2009@gmail.com

Prof. Mayank Namdev, Sagar Institute of Research & Technology (SIRT), Bhopal (Madhya Pradesh), India. E-mail: mayank.namdev@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Clustering analysis is used to find out those items that have similar characteristics and group into it. It manages the group of user information or data from Web server logs. It also can facilitate the development and execution of future marketing strategies. It dynamically support or changing a particular site based on a visitor on a return visit. An application of existing data mining algorithms, e.g. discovery of association rules or sequential patterns, the overall task is not one of simply adapting existing algorithms to new data. The WUM process is a file which having input from web user behavior as a user session files that gives an exact accounting of who accessed the website. It is also having the information just like what pages were requested and in what order, and how long each page was viewed. A user session is a time interval where a web user accesses the pages that occur during a single visit to a website. The web users access related all the information contained in a raw web server log. It does not reliably represent a user session file for a number of reasons. So that selectively information converts into tabular form and after that apply data mining technique. After getting result it produces some meaningful and useful information.

II. RELATED WORK

In 2013, Omar Zaarour, Mohamad Nagi [14] proposed an improvement of the web log mining procedure and to the prediction of online navigational pattern. It proposed for session identification using a refined time-out based heuristic. After detect the navigational pattern by using a specific density based algorithm. Now finally, a new proposed method for efficient online prediction is also recommended for applicability. In 2016, Doddegowda B J, G T Raju, Sunil Kumar S Manvi [16] having approach to personalize the information available on the Web according to user requirements. It adjusts the information/services delivered by a Web to the needs of each user or group of users to find the behavioral patterns.

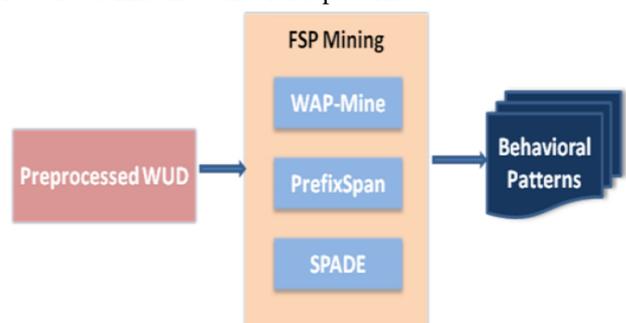


Fig. 2: System Architecture



An Analysis of User Behavior using Closed Set of Agglomerative Approach with GRC Constraints

In 2016, Minubhai Chaudhari and Chirag Mehta [17] proposed a prefixspan algorithm with GRC constraints which generates sequential patterns by using prefix projected pattern growth approach. It uses gap, compactness and recency constraints during sequential pattern mining process. The gap constraint applies limit on the separation of two consecutive transactions of discovered patterns, recency constraint makes patterns to quickly adapt the latest behaviors and compactness constraint make sure reasonable time spans for the discovered patterns.

In 2016, Fan Muhan, Shao Sujie, and Rui Lanlan [18] proposes a method for mining the frequent closed patterns in a sliding window to capture information timely and accurately when new data stream arrives. Here each basic window is used to store the Closed Pattern-tree in sliding window updates which is incrementally updated and delete the infrequent or unclosed patterns.

In 2017, H. Ryang [19] propose a novel algorithm for finding high utility patterns in the list structure over data streams on the basis of a sliding window mode. It avoid the generation of candidate patterns to improve the efficiently works in complex dynamic systems.

In 2017 Bing Zhang and Guoyan Huang [20] proposed an approach to efficiently mine sequential pattern using influential functions based on software execution sequence. It can occur multiple times in a trace, which leads to high cost of time and extreme complexity of the research.

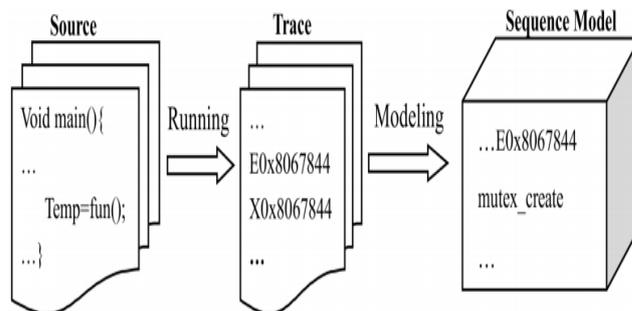


Fig. 1: Sequence Data Modeling Process

It is designed to remove repetitive patterns in software and software influential nodes mining algorithm to put forward for mine influential functions in software and to rank them by the rank-index.

III. ANALYSIS OF PREVIOUS WORK

The following table shows the analysis of previous and current work:-

S. N.	Authors	Title	Advantage	DisAdv.
1.	Bing Zhang and Guoyan Huang	Approach to mine influential functions based on software execution sequence	An effective and accurate one which combines advantages of the Page Rank and Degree based algorithms	It suffers from privilege protection loss in a release pair when it was definitely protected on all execution paths
2.	Minubhai Chaudhari and Chirag Mehta	Extension of Prefix Span Approach with GRC Constraints for Sequential Pattern Mining	It provides latest behaviors with reasonable time spans for the discovered patterns	The observed composition rules into the guessing rule set
3.	Fan Muhan, Shao Sujie, and Rui Lanlan	A Mining Algorithm for Frequent Closed Pattern on Data Stream Based on Sub Structure Compressed in Prefix-Tree	The window partitioning method to balance the time cost of mining closed patterns	The verification of their opacity without need for the original models
4.	Doddegowda B J, G T Raju, Sunil Kumar S Manvi	Extraction of Behavioral Patterns from Preprocessed Web Usage Data for Web Personalization	It provide better services in the web of each user or group of users for their behavioral patterns	The local conditional probability distribution of each node, which is calculated accordingly

IV. PROPOSED APPROACH

The proposed Profile based Closed Sequential Pattern Mining using Agglomerative Clustering approach is applied for discover frequent sequential patterns item in a cluster by using Agglomerative algorithm which produce the cluster of web data set. This cluster is used to access the partial web data set not whose web data set. So at this time this cluster data tree having the web pages of website in proportional sessions can access partially.

The following Fig. 3 shows that the process of proposed algorithm which generate useful closed sequential pattern using web data.



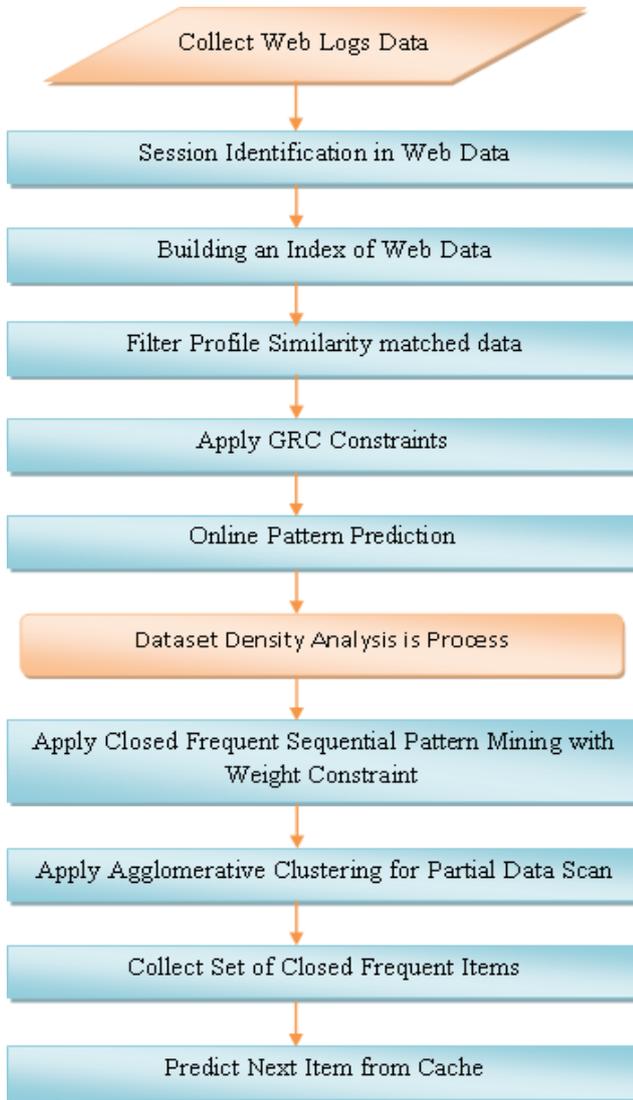


Fig. 3: The Process of Proposed Algorithm

V. CONCLUSION

In this research it scan only partial database not the whole database so that multiple scanning of database will be reduced and response time is increases. It enhanced reflection of importance of pages by using min-max weight and support of every page by using min-max weight of pages updating automatically by using web services.

It is identifying the user previous search subjects and topics so that current search will be more up to the point. So information gathered could be used to offer feedback to users on their use of the internet. It enables the effective tracking for the development and also improvement of the user interface in software by analyzing user behavior.

In future work, other data mining algorithms can be implemented in cloud to efficiency handle big data of many Hospital website in distributed environment for finding any critical diseases.

REFERENCES

1. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", in Proc. Int. Conf. Very Large Data Bases, pp. 487-499, 1994.
2. M. N. Garofalakis, R. Rastogi, K. Shim, "SPIRIT: Sequential Pattern Mining with Regular Expression Constraints", In Proceedings of

- 25th VLDB Conference, pp. 223-234, San Francisco, California, 1999.
3. M. J. Zaki, "SPADE: An Efficient Algorithm for Mining Frequent Sequences", Machine Learning Journal, Vol. 42, Issue (1-2), pp. 31-60, 2001.
4. Jian Pei, Jiawei Han and Helen Pinto, "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth", In Proceedings of 12th International Conference on Data Engineering, pp. 215-224, Heidelberg, Germany, 2001.
5. Freire J., Kumar B., and Lieuwen D., "WebViews: Accessing Personalized Web Content and Services", In Proceedings of the Tenth International World Wide Web Conference, 2001.
6. Antunes, A. L. Oliveira, "Generalization of Pattern-growth Methods for Sequential Pattern Mining with Gap Constraints", Machine Learning and Data Mining in Pattern Recognition, Third International Conference, MLDM 2003, Leipzig, Germany, July 5-7, 2003, Proceedings 2003.
7. J. Han, J. Pei, Y. Yin, and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", Data Mining Knowledge Discovery, vol. 8, no. 1, pp. 53-87, 2004.
8. J. Pei et al., "Mining Sequential Patterns by Pattern-Growth: The Prefix Span Approach", IEEE Trans. Knowledge Data Eng., vol. 16, no. 11, pp. 1424-1440, Nov. 2004.
9. Yen-Liang Chen, Ya-Han Hu, "The Consideration of Recency and Compactness in Sequential Pattern Mining", In Proceedings of the second workshop on Knowledge Economy and Electronic Commerce, Vol. 42, Iss. 2, pp. 1203-1215, 2006.
10. T.-P. Hong, C.-W. Lin, and Y.-L. Wu, "Incrementally Fast Updated Frequent Pattern Trees", Expert System Application, vol. 34, no. 4, pp. 2424-2435, May 2008.
11. Krzysztof D., Wojciech K., Marcin S., "Effective Prediction of Web User Behaviour with User-Level Models", Fundamental Informatics, IOS Press, Vol. 89, No. 2-3, pp. 189, 2008.
12. K. R. Suneetha, Dr. K. R. Krishnamoorthy, "Identifying User Behavior by Analyzing Web Server Access Log File", IJCSNS International Journal of Computer Science and Network Security, Vol. 9, No.4, pp. 327, 2009.
13. Dharendra Kumar Jha, Anil Rajput, Manmohan Singh. & Archana Tomar, (2010) "An Efficient Model for Information Gain of Sequential Pattern from Web Logs based on Dynamic Weight Constraint", IEEE International Conference on Computer Information Systems and Industrial Management.
14. Omar Zaarour, Mohamad Nagi, "Effective Web Log Mining and Online Navigational Pattern Prediction", ELSEVIER, 2013.
15. Jerry Chun, Wensheng Gan, Tzung Pei Hong, "Efficiently Maintaining the Fast Updated Sequential Pattern Trees With Sequence Deletion", IEEE Access - The Journal for Rapid open access publishing, Vol. 2, pp. 1374-1383, 2014.
16. Doddegowda B. J., G. T. Raju, Sunil Kumar, "Extraction of Behavioral Patterns from Pre processed Web Usage Data for Web Personalization", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, pp. 494-498, 2016.
17. Minubhai Chaudhari, Chirag Mehta, "Extension of Prefix Span Approach with GRC Constraints for Sequential Pattern Mining", International Conference on Electrical, Electronics, and Optimization Techniques, pp. 2496-2498, 2016.
18. Fan Muhan, Shao Sujie, Rui Lanlan, "A Mining Algorithm for Frequent Closed Pattern on Data Stream based on Sub Structure Compressed in Prefix Tree", IEEE Proceedings of CCIS, pp. 434-439, 2016.
19. H. Ryang and U. Yun., "Efficient High Utility Pattern Mining for Establishing Manufacturing Plans with Sliding Window Control", IEEE - Expert Systems with Applications, Vol. 57, pp. 214-231, 2017.
20. Bing Zhang, Guoyan Huang, Haitao He, Jiadong Ren, "Approach to Mine Influential Functions Based on Software Execution Sequence", International Journal of Engineering and Technology, Vol. 11, Issue 2, pp. 48-54, 2017.