International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 2476–2482

Analyzing User Behavior on E-Commerce Sites by Using Parallel Incremental Forward and Backward Frequent Path Traversal Approaches

Aarepu Lakshman*, Dr. B.M.G. Prasad[2], Dr. Yogesh Kumar Sharma[3]

[*]*Research Scholar, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan.*
[2]*Professor, Dept of CSE and Dean PG Courses at Holy Mary Institute of Technology and Science Hyderabad, Telangana*
[3]*Associate Professor & HOD Dep of CSE & IT, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan*

Abstract

In recent years so many e-commerce sites of various organizations are emerged around the world. The main intension behind these sites to provide essential goods. The human behavior on these sites vary continuously depends on their interests on new and attracted items which termed as web mining. As going on users may search for different items until they find desired one. To identify the user behavior we need to traverse the path associated. There exist a maximal forward and backward path traversal approach (MFPT and MBPT) which tells human behavior one-commerce sites. In existing approaches user behavior predicted up to 60-70%. To get improved user behavior analysis, in this paper we proposed a parallel incremental path traversal(PIPT) approach which uses both forward and backward frequent path traversal. With this we can analyze user behavior in accurate manner by generating a FB traversal tree. Compared to existing works our proposed approach shows better performance by means of accuracy and reduced time for analyzing user behavior.

Keywords: *web mining, MFPT, MBPT, FB-traversal tree*

1. Introduction

With the rapid expansion of internet services around the world, the number of net users has also increased as going on. This growth shows that analyzing the diverse aspects of web surfing especially web flow analysis of web mining [1][10]. And also it is very essential for website administrators and managers to analyze the behavior of various clients or users because as it provides insights hooked on the traffic coming to their sites for user needs.

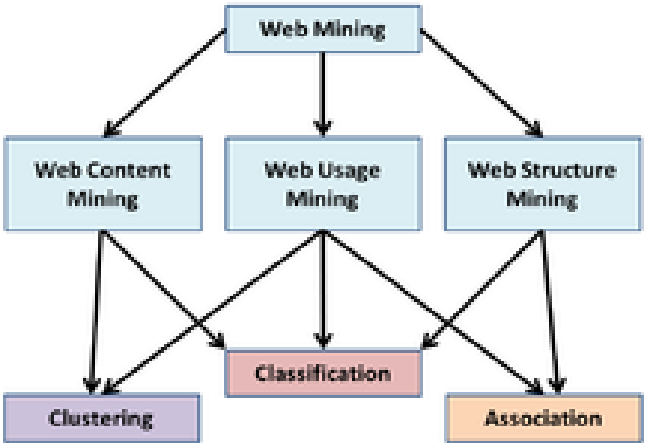

Fig 1: Types of web mining

ISSN: 2233-7857 IJFGCN

Web mining is of three types
a) Web content mining- web content mining extracts information from the webs sites, which includes text, files, images and multimedia data.
b) Web structure mining- structure mining extracts information by analyzing and modeling structure of organized data among web servers.
c) Web usage mining- It studies and analyze the usage of web pages by various users which intern reveals the interaction and relation between users and web pages.

The main role of web mining [3][4] is in E-commerce to enhance their business strategies depending on user behavior in this compete business world. Even if web content and structure mining are significant in e-commerce enlargement and administration, web usage mining ruins essential way of monitoring the advancement of the online business and represents a vibrant indication of online business. Web usage mining works on inter-page navigations of website to find out interactions of users and track sessions, and helps for better services to customers.

It might also afford a tool for adapting to online market changes, and open the door intended for e-businesses to analyze customer behaviors at E-commerce sites and blog sites. A user cycle is a complete traversal sequence for a user from entering into a E-commerce site until exit. Web Servers gathers information from E-commerce sites with the help of web logs [4] [7]. For mining frequent patterns from web logs at E-commerce sites there are different methods and approaches are there. The mined patterns are helpful the business owners to improve business, exert a pull on website design and to provide accurate surfing of items.

Web frequent Pattern mining from web logs is mostly based on association rule and concept of forward and backward references [1]. These gives a sequence of web traversal from a website which intern attain user behavior. In this paper, we go through web mining approach to discern the patterns to be formed by web site's users called as frequent surfing patterns. Maximal forward reference concept is intended to find out the paths.

The rest of this paper will focused on discussion of related work, proposed approach, results and discussions and conclusion.

2. Related Work

Website design is an important aspect for every E- commerce site to provide a flexible accessing and makes customers easy to find their needs. The navigability of E-commerce site is an important aspect in this online business world. If the users face difficulties in finding items in website then it will leads to website traffic which in turn dissatisfy the customers. This will affect the sales of particular E-commerce site to lowest growth. To overcome these issues the E-commerce site should maintain well formed structure and linking between many web pages. Some people do back track in the site more than absolutely in necessary when surfing the website. It is preferred to design a link to get back to where you are before are to start position to overcome hitting the browser back button so many times.

And it is necessary to analyze the back tracking behavior of customers with the help of traversal patterns[6] can be improve the website navigability [8]f. Traversal pattern place important role in finding the frequent pattern in users browsing and finding how website accessed by user depends on clicks. Traversal patterns reveals [2] valuable information from the user navigations to improve the structure of the website and to maintain user friendly navigation. In the information proving environment where web pages and items associated with them are linked together can help customers to traverse back and front by using icons. In some cases some items might be revisited because of locating them more times. To extract

ISSN: 2233-7857 IJFGCN

meaningful frequent patterns [9] from the log data base we can use backward traversal as well as forward traversal. Here backward references are designed to make easily travelling back but not for browsing items. This will leads to concentrate on forward references to generate sequence of traversed items. So it is important to find continuous sequence patterns and mining of these sequences [14] to analyze the user behavior on E-commerce site.

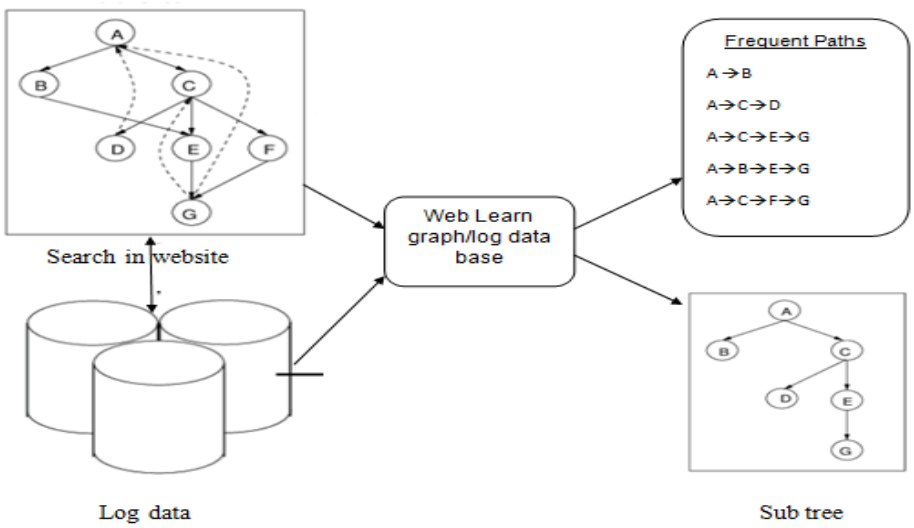

Fig 2: web mining from web track-log data

Mining of all subsequences would involve a lot of time and wearisome in sequence mining. So finding only maximal frequent patterns are probable and consumes less time. The Maximal Forward Path (MFP) as a perception of a maximal forward moving activity in visiting web documents is introduced in order to sort out the superfluous pattern from the web logs [3] [10]. The backward traversal events only happen to users in the process of searching for Web pages that really concentrate them.

Hence it is measured that only the forward browsing contains meaningful information and reflects users true browsing patterns. We must give significance to backward references also because they provides information about users navigation patterns and shows whether users are able to navigate easily within a website or not. This tells there exists a frequent backward motion which reveals that the structure of a web site is not clear.

In this paper we explore the problem of forward and backward frequent path traversal. In forward path traversal we have two phases. The first phase we convert original log data into traversal path subsequences by MF(maximal forward) references algorithm. In the second phase we call frequent traversal algorithm to find frequent traversal patterns called maximal forward references. This will generate all frequent maximal patterns [9] depending on recorded data. In the same manner maximal backward frequent path traversal approach accomplishes to get the desired backward references. Further to strengthen the prediction of user behavior we implement parallel forward and backward frequent traversal references, which targets parallel maintenance of frequent path traversals in FBPT(forward backward path traversal) tree [5].

3. Proposed Work

The proposed architecture of user behavior system consists of a tabular format of paths and path traversal tree which constructed from the tabular data. The tabular format had user ID, forward/backward

ISSN: 2233-7857 IJFGCN

paths [7], and learning paths associated with the each node of the tree. When the customer reaches the maximal forward sequences then it will stores in table from which we can draw frequent patterns. Here we consider maximal backward patterns to analyze the user behavior which is an add on to predict by following parallel way. In the proposed approach there is no generation of candidate pattern and also it scans the database only once to maintain efficiency and incrementally. Hence efficiency of proposed mining approach is better than the traditional sequential approach.

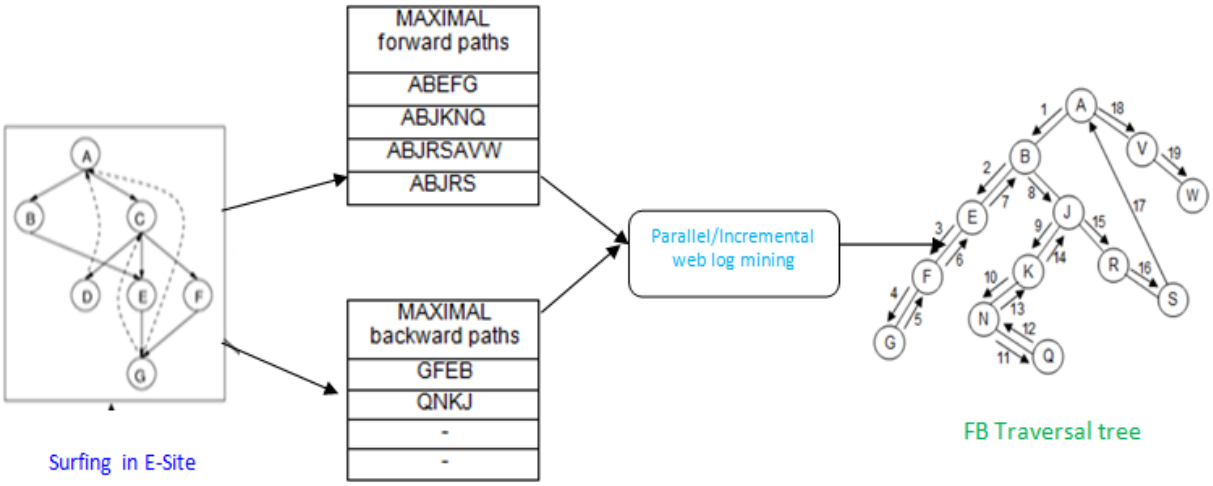

Fig 3: General user behavior in website with FB traversal tree

user ID	Learning paths	MAXIMAL forward paths	MAXIMAL backward paths
U1	ABEFG	ABEFG	GFEB
U2	ABJKNQ	ABJKNQ	QNKJ
U3	ABJRSAVW	ABJRSAVW	-
U4	ABJRS	ABJRS	-

Table 1: Learning paths with forward/backward traversals

In general the traversal log database [11] contains, for each link traversed, a pair of (source, destination). For the beginning of a new path, which is not linked to the previous traversal, the source node is null. M F is then applied to each user path to determine all of its maximal forward references. Let DF denote the database to store all the resulting maximal forward references obtained.

In e-commerce, traversal path patterns characterize the navigation behavior of customers. The information about purchasing behavior of customer can be used to find association between purchasing items and this can help in improvement of cross selling. Considering both traversal pattern and purchasing behavior of customer can add value to association rule finding. Figure 4 demonstrates traversal and purchase behavior of customer. The customer traverses and purchases items like first

ISSN: 2233-7857 IJFGCN

customer starts from A and goes to B where purchases item1. Then, customer sequentially visits E, F and G. On G, customer purchases item2

.

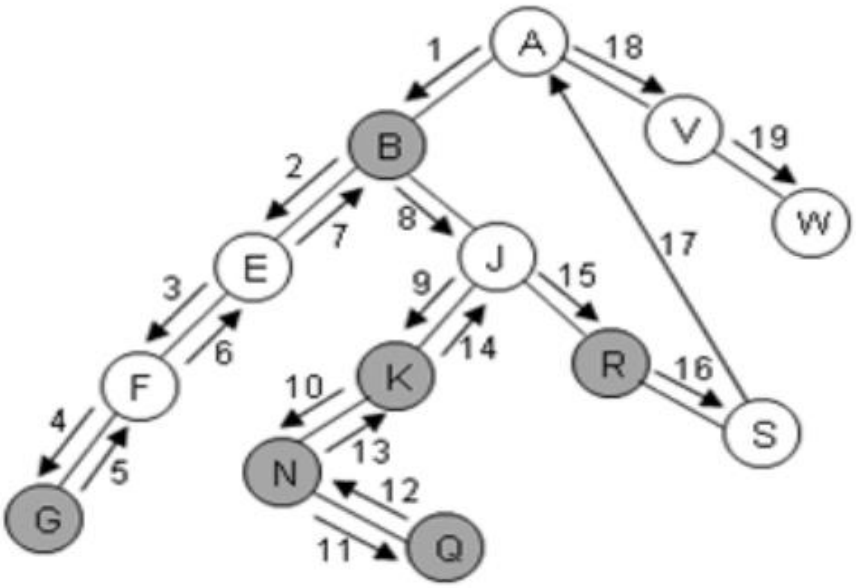

Fig 4: Traversal and purchase behavior of customer

Same way, other traversal sequences [12] are followed. The process iteratively collects all paths [13] and then generates the tree. The customer transaction detail with respect to this traversal and purchase behavior is represented in below table.

Paths	Items purchased
ABEFG	I1,I2
ABJKNQ	I1,I3,I4,I5

Table 2: Items associated with user traversal paths

The above table gives the paths and frequent items of different users. We can apply association rule mining algorithms like apriori to generate association rules which can further useful in analyzing user behavior against the frequent items. With this we can easily predicts customer behavior.

4. Results and Discussion

Experiments were made by considering reference paths with increase in path counts. Comparison of maximal forward, maximal backward and parallel incremental path traversal approaches. A PIPT is a mixed approach with both forward and backward by incremental maintenance of all frequent paths associated with the tree generated from the web logs [7]. The following table explains the comparisons of all three approaches.

ISSN: 2233-7857 IJFGCN

Table 2: Time and Item count for paths

No of Reference paths	Time(sec)			Item count		
	MFPT	MBPT	PIPT	MFPT	MBPT	PIPT
100	28	16	34	6	1	8
200	49	21	58	12	3	15
300	65	26	75	15	4	19
400	87	30	93	19	6	25

The reference path count as going on increased by 100 paths for analysis and time vary for all items will be recorded. The item count gives the total items purchased in that associated reference paths. Compared to all the proposed incremental approach gives more items which tells the customer behavior.

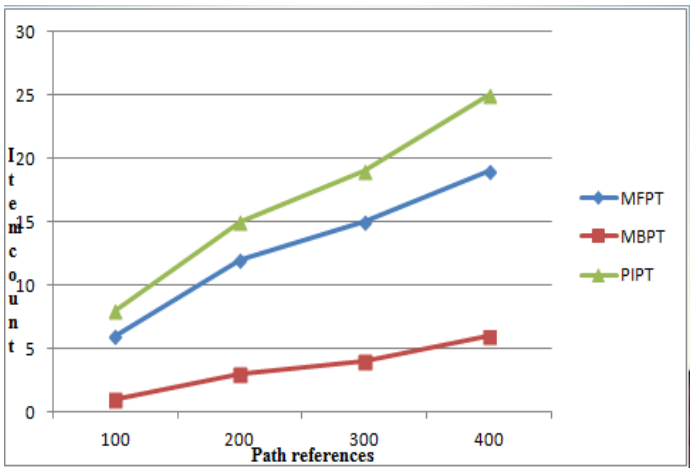

Fig 5: Path Reference count vs Items

5. Conclusion

In this paper both the maximal forward path and maximal backward path are considered and analyzed to predict the customer behavior on E-commerce site. In addition to earlier path traversal a novel parallel and incremental path traversal approach was developed to analyze the customer behavior and user interests towards items or things. Forward frequent traversal and backward frequent traversal paths reveals the user behavior up to 60-70%. So the proposed approach of parallel and incremental path traversal gives accurate results about the user behavior up to 80-90% and also insists the E-commerce owners to design a well organized and well structured website. In future we will focus on dynamic traversal approaches to predict user behavior.

References

1. Show-Jane Yen, Yue-Shi Lee, Chung-Wen Cho, “An Efficient Approach for the Maintenance of Path Traversal Patterns”, Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE’04).
2. Jatin Chhugani, Nadathur Satish, Changkyu Kim, Jason Sewall, and Pradeep Dubey, “Fast and Efficient Graph Traversal Algorithm for CPUs: Maximizing Single-Node Efficiency”, 2012 IEEE 26th International Parallel and Distributed Processing Symposium.

ISSN: 2233-7857 IJFGCN

3. Show-Jane Yen, Yue-Shi Lee and Min-Chi Hsieh, "An Efficient Incremental Algorithm for Mining Web Traversal Patterns", Proceedings of the 2005 IEEE International Conference on e-Business Engineering (ICEBE'05).
4. Zhixiang Chen, Richard H. Fowler, Ada Wai-Chee Fu_ Chunyue Wang, "Linear and Sublinear Time Algorithms for Mining Frequent Traversal Path Patterns From Very Large Web Logs", Proceedings of the Seventh International Database Engineering and Applications Symposium (IDEAS'03).
5. Yao-Te Wanga, Anthony J.T. Lee b, "Mining Web navigation patterns with a path traversal graph", Expert Systems with Applications 38 (2011) 7112–7122
6. Lin Zhou, Ying Liu, Jing Wang, Yong Shi, "Utility-based Web Path Traversal Pattern Mining", Seventh IEEE International Conference on Data Mining – Workshops.
7. Jian Pei, Jiawei Han, Behzad Mortazavi-asl, Hua Zhu, "Mining Access Patterns Efficiently from Web Logs".
8. Ajith Abraham, Vitorino Ramos, "Web Usage Mining Using Artificial Ant Colony Clustering and Genetic Programming".
9. Arthur.A.Shaw, N.P. Gopalan, "Frequent Pattern Mining of Trajectory Coordinates using Apriori Algorithm", International Journal of Computer Applications (0975 – 8887) Volume 22– No.9, May 2011.
10. Federico Michele Facca, Pier Luca Lanzi, "Mining interesting knowledge from weblogs: a survey", Data & Knowledge Engineering 53 (2005) 225–24.
11. C.I. EZEIFE, YI LU, "Mining Web Log Sequential Patterns with Position Coded Pre-Order Linked WAP-Tree", Data Mining and Knowledge Discovery, 10, 5–38, 2005 Springer Science + Business Media, Inc.
12. Rakesh Agrawal, Ramakrishnan Srikant, " Mining Sequential Patterns".
13. Anna Gutowska, Luis L. Perez, " A Comparison of Methods for Classification and Prediction of Web Access Patterns", COMP540 - Final Report - Spring 2010.
14. R. Agrawal, C. Faloutsos, and A. Swami. Ef-cient Similarity Search in Sequence Databases.Proceedings of the 4th Intl. conf. on Foundations of Data Organization and Algorithms, October, 1993.

ISSN: 2233-7857 IJFGCN