ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

Comparative Study of Web Mining Algorithms for Web Page Prediction in Recommendation System

Rahul Neve [1], K.P Adhiya [2]

M.E Scholar, Dept of Computer Science and Engineering, SSBT's COET, Bambhori - Jalgaon, India[1]
Associate Professor & Head , Dept of Computer Engineering, SSBT's COET,Bambhori - Jalgaon, India[2]

ABSTRACT: **This paper shows the comparative study and implementation of recommendation system based on two different web mining algorithms. The proposed system is designed for web page prediction in recommendation system as well as it is helpful for the study of web mining algorithm to get frequent sequential access pattern from web log file of web server. The experiments are conducted based on the implementation of the algorithms (GSP and Prefix Span) and results are noted down in which it is found that pattern growth based algorithm(Prefix Span) is efficient than apriori based algorithm (GSP) for web log sequential access pattern mining.**

Keywords: Web Mining, Sequential access pattern, Data Mining, Web usage mining.

I. INTRODUCTION

The ample amount of information available on World Wide Web, which lacks an integrated structure or schema; it becomes much more difficult for users to access relevant information efficiently. Analysing and modelling web navigation behaviour is helpful in understanding demands of online users. Following that, the analyzed results can be seen as knowledge to be used in intelligent online applications, refining website maps, and web based personalization system and improving searching accuracy when seeking information. The main goal of the recommendation system is to improve Web site usability. Typically, the Web usage mining prediction process is structured according to two components performed online and off-line with respect to Web server activity.

The World Wide Web serves as a huge, widely distributed, global information service centre for news, advertisements, consumer information, financial management, education, government, e-commerce and many other information services. There are many approaches for mining frequently occurring patterns. Sequential access pattern mining [2] is one the latest techniques for web usage mining.

II. LITERATURE SURVEY

Sequential pattern mining is the mining of the frequently occurring ordered events or subsequences as patterns. An example of sequential pattern is "*Customers who buy a notebook are likely to buy pen or pencil*". The Sequential mining pattern was first introduced by *Agrawal* and *Srikant* in 1995 based on their study of customer purchase sequence. Sequential pattern can be applied includes Web access pattern analysis.*Agrawal*, define a problem of finding inter-transaction patterns. Under this context, a pattern is an ordered list of sets of items. In sequential pattern, frequently occurring event sequences (temporal patterns) in the data are identified. This technique is very useful for the identification of navigational patterns. Sequential patterns can be discovered by using two methods, by using deterministic techniques (recording the navigational behavior of the user) and stochastic methods (use the sequence of web pages that have been visited in order to predict subsequent visits). The data source for sequential item or events is web log.[1]

Chen in 1996 introduced the concept of "*maximal forward references"*.[7] The *Maximal Forward (MF) references* algorithm aims at converting the original sequence of Web server logs into a set of traversal patterns based on statistically dominant paths and association rules. In an information-providing environment where objects are linked together, users are apt to traverse objects back and forth in accordance with the links and icons provided. As a result, some node might be revisited because of its location, rather than its content. For example, in a WWW environment, to reach a sibling node a user is usually inclined to use "backward" icon and then a forward selection, instead of opening a new URL. Consequently, to extract meaningful user access patterns from the original log database, we naturally want to take into consideration the effect of such backward traversals and discover the real access patterns of interest. a backward reference is mainly made for ease of traveling but not for browsing, and concentrate on the discovery of forward reference patterns. Specifically, a backward reference means revisiting a previously visited object by the same user access. When

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

backward references occur, a forward reference path terminates. Therefore it is known as *maximal forward reference*. MF references can be defined as the sequence of documents requested by a user up to the last one before backtracking. The approach proposed works in two steps. The first one is the MF algorithm. In the second step they proposed two more algorithms. One determines the large reference sequences using hashing and pruning techniques and the other further improves it in terms of the database scans. The selective scan option is the strength of their technique.

In year 1996 GSP Algorithm was introduce by Srikant and Agrawal for sequential access pattern mining. The GSP Algorithm makes multiple passes over data. The first pass determines the frequent 1-item patterns ($L1$). Each subsequent pass starts with a seed set: the frequent sequences found in the previous pass ($Lk-1$). The seed set is used to generate new potentially frequent sequences, called candidate sequences (Ck)[5]. Each candidate sequence has one more item than a seed sequence. In order to obtain k-sequence candidate Ck , the frequent sequence $Lk-1$ joins with itself Apriori-gen way. The algorithm terminates when there are no frequent sequences at the end of a pass, or when there are no candidate sequences generated. The GSP algorithm uses a hash tree to reduce the number of candidates that are checked for support in the database.[1]

Mortazaviasl introduced the idea of progressively partitioning the database of user sequences into smaller sub databases for mining sequential patterns. They introduced a new projection-based algorithm for mining sequential patterns called the *PrefixSpan*. It uses the frequent sequence trellis to partition the database. It traverses the frequent item lattice level-by-level in depth-first order. PrefixSpan exhibit a greater selectivity while in the FreeSpan it is not guaranteed that a projected database would shrink in size or not. Because of changing business needs some sequential patterns become invalid and some needs to be updated which are not solved with Prefix Span. The PSP approach is much similar to the GSP algorithm (Srikant and Agrawal, 1996). At each step k, the database is browsed for counting the support of current candidates. Then, the frequent sequence set, Lk is built..[3]

The next development is *FreeSpan*(i.e **Fre**qu**e**nt Pattern Projected **S**equential **Pa**ttern Mining) [9] .Its general idea is to use the frequent items to recursively project sequence databases into set of smaller databases and grow subsequence fragment in each projected databases. This process portion both the data and set of frequent patterns to be tested and confined each test being conducted to corresponding smaller projected databases. FreeSpan uses divide and conquer method to find complete set of sequential pattern. Then a *frequent item matrix* F is constructed to count the occurrence of frequency of each length. This frequent item matrix is used to generate the length – *n sequential pattern.* Hence for a given sequence database *S* and the minimum support threshold *freespan* mines the complete set of sequential pattern.

The *FreeSpan* projects a large sequence database recursively into a set of small projected sequence database based on currently mined frequent sets[11]. The sub-sequent mining is confined to each projected database, relevant to smaller set of candidates. The alternate-level in FreeSpan reduces the cost of scanning multiple projected databases when finding frequent length greater than 2 candidate filtering, therefore it reduce the effort of repeatedly generating and checking the large set of candidate sequences against the entire database and achieve better performance than Apriori-Like GSP algorithm.

Pei (in 2000) introduced *WAP-mine* algorithm [8]in which mining is performed in a Web access sequence database to depict access patterns. The length of web log pieces can be very long in web mining and it can be in a huge number. They use a conditional search strategy and a proposed *WAP-tree structure.* The WAP-tree stores complex and critical Information for web access pattern mining. It stores all count information for pattern mining and thus makes mining process free from counting candidates by pattern matching. The search strategy WAP-mine is proposed to mine non-redundant web access patterns.

Although WAP-tree algorithm scans the original database only twice and avoids the problem of generating explosive candidate sets as in Apriori-like algorithm, its main drawback is recursively re-constructing large numbers of intermediate WAP-trees and patterns during mining taking up computing resources. Pre-Order linked WAP tree algorithm (PLWAP) [5][6] is a version of the WAP tree algorithm that assigns unique binary position code to each tree node and performs the header node linkages pre-order fashion (root, left, right).Both the pre-order linkage and binary position codes enable the PLWAP to directly mine the sequential patterns from the one initial WAP tree starting with prefix sequence, without re-constructing the intermediate WAP trees. To assign position codes to a PLWAP node, the root has null code, and the leftmost child of any parent node has a code that appends '1' to the position code of its parent, while the position code of any other node has '0' appended to the position code of its nearest left sibling. The PLWAP technique presents a much better performance than that achieved by the WAP-tree technique as shown in extensive performance analysis. PLWAP algorithm uses a preorder linked, position coded version of WAP tree and eliminates the need to recursively re-construct intermediate WAP

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

trees during sequential mining as done by WAP tree technique. PLWAP produces significant reduction in response time achieved by the WAP algorithm and provides a position code mechanism for remembering the stored database, thus, eliminating the need to re-scan the original database as would be necessary for applications like those incrementally maintaining mined frequent patterns, performing stream or dynamic mining.[10]

III. EXISTING METHODOLOGY

A. Generalized Sequential Patterns (GSP):

It uses the downwards-closure property of sequential patterns and adopts a multiple-pass, candidate generate-and-test approach. This algorithm is outlined as follows:

In the first scan of database, it finds all of the frequent items > = minimum support. Each such item yields a 1-event frequent sequence consisting of that item. Each subsequent pass starts with seed set of sequential patterns-the set of sequential pattern found in previous pass. This seed is used to generate new potentially frequent patterns, called candidate sequences. Each candidate sequence contains one more item then the seed sequence pattern from which it was generated (where each event in the pattern may contain one or multiple items). The number of instances of items in a sequence is the length of the sequence. So, all of the candidate sequences in the given pass will have the same length.

The algorithm terminates when no new sequential pattern is found in a pass, or no candidate sequence can be generated.

1) Example for GSP Algorithm: Assume below web access sequence database

TABLE I:
WEB ACCESS SEQUENCE DATABASE

UID	Web access Sequence
100	abdac
200	eaebcac
300	babfaed
400	afbacfc

Consider WASD as above. Let the minimum support threshold is equal to 3. Hence the frequent item is one whose occurrence is more than or equal to minimum support in *WASD*. Following are the steps to count frequency of item in *WASD*.

The frequency of given items in WASD is a:4 , b:4, c:3 , d: 2, e:2 , f:2. Frequent item set from above is {a: 4, b: 4, c: 3}, as these are more than or equal to minimum support.

As the frequent items are { a:4, b:4, c:3} , hence 1-length sequence is { a,b,c}.Now for length-2 sequence candidate set is generated by joining process as {aa, ab,ac,ba,bb,bc,ca,cb,cc}.After this pruning is done on the candidate sequence of length-2 {aa:4, ab:4,ac:3,ba:4,bb:1,bc:3,ca:1,cb:0;cc:2} ,the result of pruning for length-2 is as follows {aa,ab,ac,ba,bc}. Now the seed for the next sequence is { aa,ab,ac,ba,bc}.This seed generate next candidate set sequence. {aaa,aab,aac,aba,abb,abc,aca,acb,acc,baa,bab,bac,bca,bcb, bcc}.
{aaa:0,aab:0,aac:3,aba:4,abb:0,abc:3,aca:1,acb:0,acc:2,baa;1,bab:1,bac:3,bca:1,bcb:0, bcc:2}.

TABLE I
GSP ALGORITHM RESULTS FOR GIVEN EXAMPLE

Candidate Set	Frequency Count	Pruning Result
{ a,b,c}	{ a:4, b:4, c:3}	{ a,b,c}
{aa, ab,ac,ba,bb,bc ,ca,cb,cc}	{aa:4,ab:4,ac:3,ba:4, bb:1,bc:3,ca:1, cb:0;cc:2}	{aa,ab,ac,ba, bc}
{aaa,aab,aac,a ba,abb,abc, aca,acb,acc,ba a,bab,bac,bca, bcb,bcc}.	{aaa:0,aab:0,aac:3,a ba:4,abb:0, abc:3, aca:1,acb:0,acc:2,ba a;1,bab:1,bac:3,bca: 1,bcb:0,bcc;2}.	{ aac, aba, abc, bac}
{aaca,aacb,aa cc,abaa,abab, abac,abca,abc b,abcc,baca, bacb,bacc}	{aaca:0,aacb:0,aacc: 1,abaa:0,abab:0;abac :3,abca:1,abcb:0,abc c:2,baca:0, bacb:0,bacc:1}	{abac}
{abaca,abacb, abacc}	{abaca:0,abacb:0,ab acc:0}	At this level pruning stop and algo terminates

B. Prefix Span Algorithm:

Prefix Span comes under pattern growth method for mining sequential patterns. Its major idea is that, instead of projecting sequence databases by considering all the possible occurrences of frequent subsequence's, the projection is based only on frequent prefixes because any

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

frequent subsequence can always be found by growing a frequent prefix. Prefix Span examines only the prefix subsequence's and projects only their corresponding postfix subsequence's into projected databases. This way, sequential patterns are grown in each projected database by exploring only local frequent sequences. Three major steps of prefix span are as follows:

1. Find all 1-itemset sequential patterns by scanning the database WASD.

2. Divide the search space to get projected databases: The complete set of sequential patterns can be partitioned into the following three subsets according to the three prefixes: i) the ones with prefix ***a***, ii) the ones with prefix ***b*** iii) the ones with prefix ***c.***

3. Find subset of sequential patterns: The subsets of sequential patterns can be mined by constructing the corresponding set of projected databases and mining each recursively. The projected databases as well as sequential patterns found in them are listed in Table III,

TABLE III:
Projected Sequence database and final respective sequential pattern of prefix span

Prefix	**Projected sequence database**	**Sequential patterns**
a	{bac ,bcac,ba,bacc}	{a:4} { aa:4,ab:4,ac :3} {aac:3,aba;3,abc:4,} {abac:4}
b	{ac , cac,a,acc}	{b:4} {ba:4,bc:3} {bac:3}
c	{_,ac,c}	{c;3}

IV PROPOSED WORK AND IMPLEMENTATION

The proposed web page prediction recommend system is developed using GSP and Prefix Span algorithm. The goal of the recommender system is to determine which web pages are more likely to be accessed next page by the current user in the near future. The proposed model is shown figure 1.

Figure 1 shows the architecture of the recommender system. First, all users' web access activities of a website are recorded by the Web server of the website and stored into the Web Server Logs. Web Usage Mining (WUM) is the application of data mining techniques to discover usage patterns from Web data. In a general process of WUM, distinguish three main steps: data pre-processing, pattern discovery and pattern analysis.

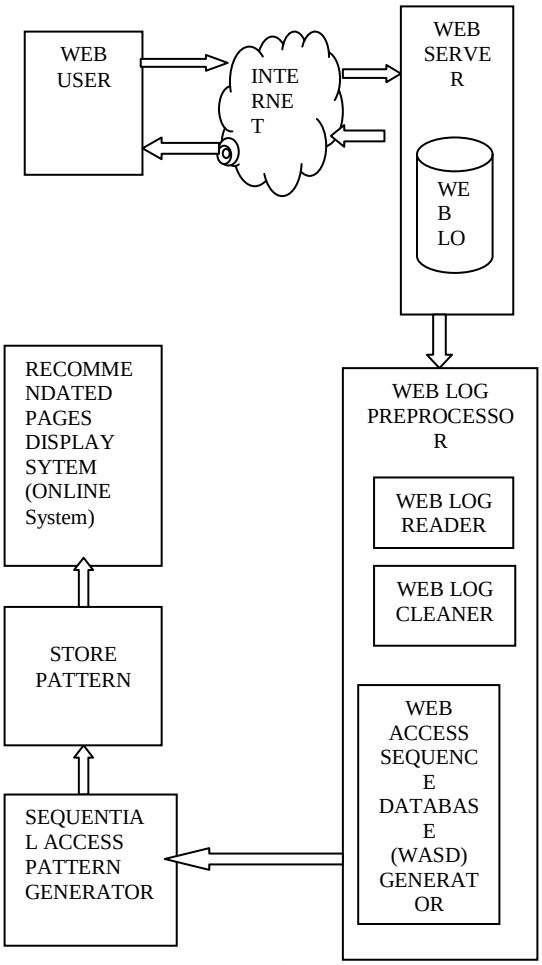

Fig 1: Architecture of Web Recommender System

The proposed model consist of four main modules

A. Web Log Pre-processor

Web log Reader: Web log reader helps to read the web log text file from the specified path. During preprocessing phase, raw Web logs need to be cleaned, analyzed and converted before further pattern mining. The data recorded in server logs, such as the user IP address, browser, viewing time, etc, are available to identify users and sessions. However, because some page views may be cached by the user browser or by a proxy server, we should know that the data collected by server logs are not entirely reliable. This problem can be partly solved by using. some other kinds of usage information such as

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

cookies. After each user has been identified, the entry for each user must be divided into sessions. A timeout is often used to break the entry into sessions.

Fig 2: Web Log Reader

Example of web log file:

199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
unicomp6.unicomp.net - - [01/Jul/1995:00:00:06 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
199.120.110.21 - - [01/Jul/1995:00:00:09 -0400] "GET /shuttle/missions/sts-73/mission-sts-73.html HTTP/1.0" 200 4085
burger.letters.com - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/countdown/liftoff.html HTTP/1.0" 304 0
199.120.110.21 - - [01/Jul/1995:00:00:11 -0400] "GET /shuttle/missions/sts-73/sts-73-patch-small.gif HTTP/1.0" 200 4179

Web log Cleaner: During field extraction which the related fields from the web log txt file is selected and only these filed are kept which will be useful for the mining process. In this proposed system the field that are being selected for mining are *IP-address, data-time, page url*

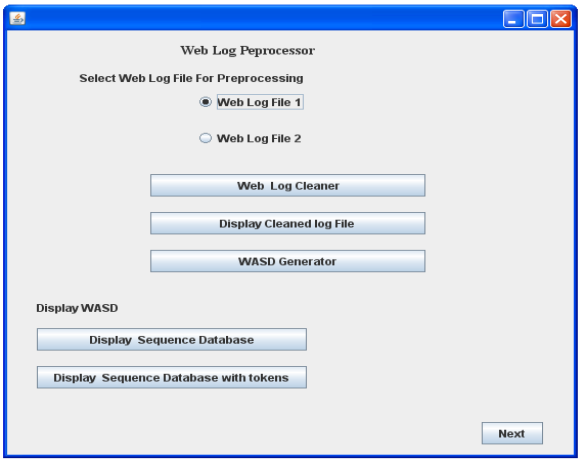

Fig 3: Web Log Prepossor(Web Log Cleaner and WASD Genrator)

Web access Sequence Database Generator (WASD): WASD Generator takes input from the Web log Cleaner module (i.e. Cleaned web log file). It contains IP-address or domain name followed by date time filed and *URL's* of the web page.

- List of Unique *URL's* are also collected in array List.

Example Of unique *URL's*:
shuttle/missions/sts-38/mission-sts-38.html
/shuttle/countdown/launch-team.html
/history/skylab/skylab-4.html
/msfc/onboard/onboard.html
/shuttle/missions/sts-5/sts-5-info.html
/shuttle/technology/sts-newsref/sts-apu.html

- Token is provided to each Url in the List.

Example: Unique *URL's* with tokens
/shuttle/missions/sts-38/mission-sts-38.html 1
/shuttle/countdown/launch-team.html 2
/history/skylab/skylab-4.html 3
/msfc/onboard/onboard.html 4
/shuttle/missions/sts-5/sts-5-info.html 5
/shuttle/technology/sts-newsref/sts-apu.html 6

- The grouping process of IP address / domain name is done. In front of each IP address /domain name the sequence of *URL's* are mentioned according to the access log information.

Example: IP address or domain name is followed by the sequence of *URL's*
199.120.110.21 /shuttle/missions/sts-73/mission-sts-73.html /shuttle/missions/sts-73/mission-sts-73.html

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

burger.letters.com
/shuttle/countdown/liftoff.html
/shuttle/countdown/liftoff.html
/shuttle/countdown/liftoff.html
/shuttle/countdown/liftoff.html

- At the end of process the IP address/domain name are followed by the sequence of tokens of *URL's* are mentioned.

Example: WASD
199.120.110.21 128?128?
burger.letters.com
69?69?69?69?69?69?69?69?69?69?69?69?69?
205.212.115.106 80?59?80?59?

B. Sequential Access Pattern Generator

This module of proposed system is used to generate the frequent access pattern of web pages from the WASD. Input to this module is WASD (IP-address with token of *URL's* sequences as shown in figure).Web mining algorithms are implemented in this module.

This module consists of two algorithms (GSP and Prefix Span), which are described detailed in chapter three of report. This module is also helpful for the comparative study of these two algorithms in detail.

Algorithm GSP(S)
Step 1 Scan WASD for first sequence that is initial-pass(*S*); the first pass over *S.Get initial pass in Candidate*
Step 2 $F1 \leftarrow \{\langle\{f\}\rangle | f$ belongs $C1, f.count/n \geq minsup\}$; is the number of sequences in *S*
Step 3 for ($k = 2$; $Fk-1$is not empty k++) subsequent passes over *S*
Step 4 $Ck \leftarrow$ candidate-gen-SPM($Fk-1$);
Step 5 for each data sequence *s* belongs to *S* scan the data
Step 6 for each candidate *c belongs to Ck* do
Step 7 if *c* is contained in *s* then
Step 8 increment the support count
Step 9 endfor
Step 10 endfor
Step 11 $Fk \leftarrow \{c$ belongs to $Ck \mid c.count/n \geq minsup\}$
Step 12 endfor
Step 13 return *Fk*;

Algorithm Prefix Span
Step 1: Scan the WASD once
Step 2: Get the frequent item from the WASD such that the occurrence of *fi* should be equal to or greater than minimum support
Step 3: For i= 0 to list of all frequent item
  i) On basis of frequent item list get the suffix subsequences from WASD of that supplied *fi[i]*.
  ii) Get the frequent item from the subsequence and append to the List
Step 4: Repeat Step 3 till end of all frequent items

C. Store Pattern

This module of the system is used to convert the frequent access pattern which is in form of tokens into the *URL's* sequence. This module help to generate result text file of frequent pattern which is further used by the Online web page recommendation system to display the final result to the user.

D. Recommendation Pages Display System.

It is the interface given to user in form of web site to know the recommended web pages of that user on basis of mining algorithm.

V RESULT AND DISCUSSION

For the study of web mining algorithms a system is developed which incorporate two different algorithms 1) GSP and 2) Prefix Span. The experiments were conducted on 1GHZ AMD PC with 2GB of main memory. The *dataset* of *NASA access log was used.*

Experimental was conducted on NASA access log dataset of 1Mb containing 10000 records in raw log file. After cleaning process on that file the size of cleaned file was 171kb containing 1986 numbers of record. Unique IP address was 645 and unique *URL's* were 278.Following table shows the experimental result of GSP and Prefix Span with minimum varying minimum threshold. The performance evaluation of algorithm for execution time is shown in below table

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

TABLE IV
Experimental Execution time of GSP and Algorithm by varying minimum support

Minimum Support Threshold (Frequency of web pages)	**GSP Algorithm Execution time (in ms)**	**Prefix Span Algorithm Execution time (in ms)**
3	7313	218
6	2359	125
9	1047	94
12	562	87
15	438	78
18	328	31
21	234	15

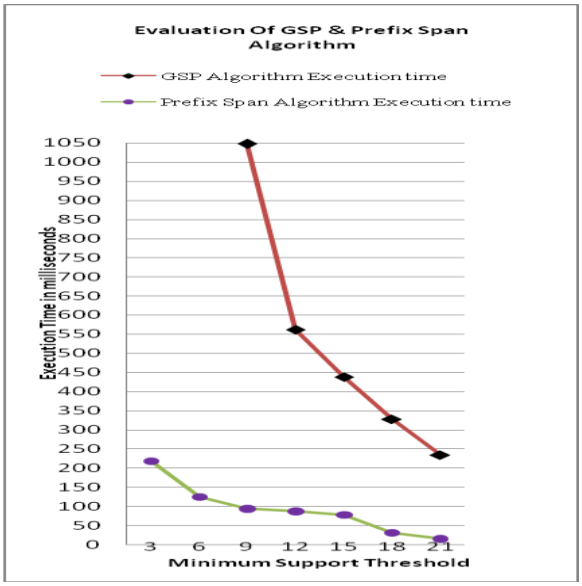

Figure : 4 Performance evaluation graph for GSP and Prefix Span

Above evaluation is based on the experimental result of GSP and prefix Span algorithm. It shows the execution time of both algorithms after varying the minimum support count. It is clearly seen that for low minimum support there is large difference between execution time. Prefix Span algorithm executes much faster than GSP for the low minimum support. As the minimum support increases the difference between the times of execute of both algorithm reduces.

VI CONCLUSION AND FUTURE SCOPE

Web mining algorithms are used to get the frequent sequential access pattern from the dataset such as web log. Web log cleaning is the necessary step for mining the web log data. The implemented system consists of two web mining algorithms for study of sequential access mining. The system is developed for the online recommendation of web pages.

The comparative study of two algorithms is done and following are the conclusion

TABLE V
CONCLUSION

Parameter	**Algorithm**	
	GSP	**Prefix Span**
Category	Apriori – based	Pattern Growth based
Approach	Generate and test approach for mining frequent pattern	Divide and conquer approach for mining frequent pattern
Scanning of Database	Original database is scanned multiple time	Original database is scanned only once.
Candidate set	For mining each time candidate sets are generated	No need of candidate set, instead of that the projected sequence database at run-time is generated.
Partitioning	Partitioning of the search space is not reqired.	Search space partitioning is applied based on frequent items or frequent sub

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

		sequences
Execution time for low minimum support and dense database	Very High as it is proved from the experimental result shown in table(7313 ms for 1 Mb of web log file)	Comparatively very low as shown in the experimental result in table (218 ms for 1 Mb of web log file)

In future the web access patterns can be analyzed for semantic web, this can include deeper information: what is the meaning of the pattern; what are the synonym patterns; and what are the typical transactions that this pattern resides? In many cases, frequent patterns are mined from certain data sets which also contain structural information For example, the shopping transaction data is normally tagged with time and location. Some text data (e.g., research papers) has associated attributes (e.g. authors, journals, or conferences). A contextual analysis of frequent patterns over the structural information can help respond questions like "why this pattern is frequent?" An example answer could be "this pattern is frequent because it happens heavily during time *T*1 to *T*2".

REFERENCES

[1] Jiawei Han and Micheline Kamber ,"*Data Mining Concepts and Techniques*",second edition ,Morgan Kaufmann publisher.
[2] Xiaogang Wang, Yan Bai "An Information Retrieval Method Based On Sequential Access Patterns" 978-0-7695-4003-2010 IEEE 2010 Asia-Pacific Conference on Wearable Computing Systems.
[3] Nasser Ahmed Sajid,Salman Zafar "Sequential Pattern Finding :A Survey" 978-1-4244-8003 , 2010 IEEE
[4] Baoyao Zhou ,Siu Cheung Hui "Efficient sequential access pattern mining for web"recommendations" International Journal of Knowledge-based and Intelligent Engineering Systems ,Volume 10 Issue 2 ,April 2006
[5] C.I. Ezeife "PLWAP Sequential Mining: Open Source Code" – August 21,2005,Chicago, Illinois, USA.ACM.
[6] Ezeife and Y. Lu. " Mining web log sequential patterns with position coded pre-order linked wap-tree. " International Journal of Data Mining and Knowledge Discovery (DMKD) Kluwer Publishers, 10(1):5–38, 2005.
[7] Ming-Syan Chen, Jong Soo Park "Efficient Data Mining for Path Traversal Patterns" 1041-4347/98, 1998 IEEE
[8] Jian Pei,Jiawei Han, "Mining Access Patterns Efficiently from Web Logs" supported in part by Natural Science and Engineering Research Council Of Canada (grant NSERC-A3723).(Un-Cited).
[9] Jiawei Han,Jian Pei, "FreeSpan:Frequent Pattern-Projected Sequential Pattern Minig" supported in part by Natural Science and Engineering Research Council Of Canada (grant NSERC-A3723).(Un-Cited).
[10] S. Vijayalakshmi and V. Mohan "Mining Sequential Access Pattern with Low Support from Large Pre-Processed Web Logs" Journal of Computer Science 6 (11): 1293-1300, 2010 ISSN 1549-3636
[11] Jiawei Han,Jian Pei, "FreeSpan:Frequent Pattern-Projected Sequential Pattern Minig" supported in part by Natural Science and Engineering Research Council Of Canada (grant NSERC-A3723).

Biography

Mr. K P Adhiya , Associate Professor and Head Of Department of Computers Engineering in SSBT's College of Engineering and Technology Bambhori Jalgaon, Maharastra, 21 Years of Experience in teaching ,Pursuing PhD in Engineering

Mr. Rahul P Neve ,M.E Scholar ,Pursuing M.E in Computer Science and Engineering from North Maharastra University