



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 8, August 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

An Improved Method of User Behavior Detection Employing Region Specific Search

Meenakshi Pipariya¹ Ashish Gupta²

¹M. Tech Student, Department of Computer Science & Engineering, NITM, Gwalior, India

²Assistant Professor, Department of Computer Science & Engineering, NITM, Gwalior, India

ABSTRACT: The development of the web in past few years has created a lot of challenge in this field. The new work in this field is the search of the data in a search tree pattern based on tree. Various sequential mining algorithms have been developed till date. Web usage mining is used to operate the web server logs, that contains the navigation history of the user. Recommendation system is explained properly with the explanation of whole procedure of the recommendation system. The search results of the data leads to the proper and efficient search. But the problem was the time utilization and the search results generated from them. So, a new local search algorithm is proposed for country-wise search that makes the searching more efficient on local results basis. This approach has led to an advancement in the search based methods and the results generated.

KEYWORDS: Sequential pattern mining, web mining, recommendation system

I. INTRODUCTION

Data mining is an extension of traditional data analysis and statistical approaches. In addition it integrates analytical techniques taken from a range of disciplines including, but that is not limited to pattern matching numerical analysis, and areas of artificial intelligence such as neural networks, genetic algorithms and machine learning, [1].

The Internet has grown into a gold mine that comprises or dynamically generates information beneficial to E-businesses. With the help of rapid growth in sources of information available on the World Wide Web (www) and the rapidly growing pace of adoption to Internet commerce [2].

Web mining can be classified as the discovery and analysis of useful information derived from the World Wide Web (WWW), which is based on different ways to gather information and different emphasis. Web Usage Mining can be determined as the discovery and analysis of user access patterns, by associated data from a specific Web site and mining of log files.

Web usage mining can be accomplished first by reporting visitors' traffic information based on Web server log files and other source of traffic data (as discussed below). Web server log files were utilized initially by the system administrators and webmasters for the purposes of "what kind of errors are being generated, how much traffic they are getting and how many requests fail", etc. Though, Web server log files can also trace and record the visitor's on-line behavioral aspects. For example, after some basic traffic analysis, the log files can help us to answer questions such as "What pages are the most and least popular? from which search engine are visitors coming?" [3]. Web log files are one way to gather Web traffic data. The another way is to "sniff" TCP/IP packets as they cross the network, and to "plug in" to each Web server. After the Web traffic data is gained, it may be combined with other relational databases, over which the implementation of data mining techniques can be done.

Recommendation system is a subclass of filtering information system that pursue to predict the 'rating' or 'preference' that user would give to an item. Recommendation systems have become extremely common in recent years, and can be applied in a variety of applications. The most popular ones are probably music, books, news, research articles, movies, music, news, books, research articles, search queries, social tags, and products in general.

II. WEB MINING

Web can be explained as a complex system with interconnected elements and mining is the process to extract information or data, so web mining is the process of data extraction from this interconnected system. Web mining can be separated into three different types – Web usage mining, Web content mining and Web structure mining [4].

A. Web Content Mining

It refers to mine the information from weblog or record set. In this technique content like picture text and graphs are scanned to find out the significance of the content from the requested query. Web page Content mining is a

traditional search process of web page via content [5].

B. Web Structure Mining

This technique refers to mine the information from the linkage structure of the Webpage then this linkage information is used to capture the list of interesting patterns. Web Structure mining mines the structures like HTML or XML tags and extracts information from the actual organization of the pages. To give weight to the page, it uses interconnections between web pages [5].

C. Web Usage Mining

This is the application of data mining techniques. Web Usage mining refers to discover interesting usage patterns from Web data in order to understand and better serve the requirements of Web-based applications. Web usage mining refers to mine the information gathered from the web access behavior of the customers at the same time as they are investigating and skimming through the Web[6]. Web usage mining is the application of data mining techniques to appreciate the web usage patterns. It mines data from logs, cookies, bookmarks, , scrolls, mouse click, user sessions, user profiles, user queries etc. The three phases of webusage mining are preprocessing, pattern discovery, and pattern analysis [6].

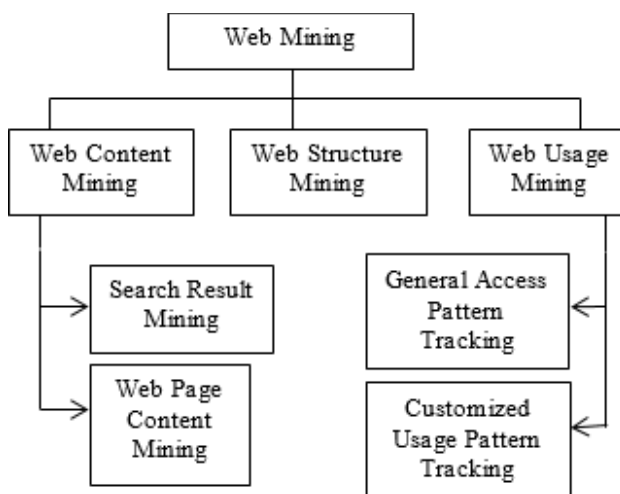


Fig. 1: Web Mining Classification

Sequential pattern is a sequence of itemsets that frequently occurred in a specific order, all items in the same itemsets are supposed to have the same transaction timevalue or within a time gap.

III. PROCEDURE OF RECOMMENDATION SYSTEM

Recommender systems now became famous commercially and in the research community, where many approaches have been recommended for providing recommendations. In numerous cases a system designer desires to employ a recommendation system must choose between a set of candidate approaches. A first step towards choosing an appropriate algorithm is to resolve which properties of the application need focus upon when making this choice. Indeed, recommendation systems have a variety of properties that may affect user experience such as scalability, accuracy, robustness, and so forth. Recommender systems are having the effect of guiding users in a personalized way as the useful and interesting objects in a large space of possible options [7]. As the input Recommendation algorithms are using customer’s interests to generate a list of recommended items. Amazon.com, uses recommendation algorithms to personalize the online store for each n every customer [8]. Recommending items related issues has been studied extensively and mainly the two paradigms have emerged. First is Content-based recommendation systems that tries to identify recommend items similar to those a given user has liked before in the past, whereas the second systems designed recommendation systems, according to the collaborative recommendation paradigm identifies those users whose preferences are similar to the given user and recommend items they have liked [9].

A. Types of Recommendation System:

1) Content-Based Recommendation System:

The system learns to those recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items. For example, if a user has positively rated a movie that belongs to the comedy genre, then the system can learn to recommend other movies from this genre. The overview of content based recommender systems, imposing some order among the extensive and diverse features involved in their design and implementation. Content-based Recommendation system discusses trends and the future research which might lead towards the next generation of recommender systems. Collaborative filtering: The simplest and original implementation of this approach [10] recommends to the active user the items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating history of the users. This is the reason why [11] refers to collaborative filtering as “people-to-people correlation.” Recommendation procedure is partitioned into following steps:-

- 1) Collect Web logs
- 2) Apply pre-processing
- 3) Prepare processed data base
- 4) Apply Pattern mining algorithm
- 5) Construct Pattern tree
- 6) Recommendation rule generation

a) Web logs:

First we will collect Web logs (Web server access Log) of any Web-site. Web logs are the text file (ASCII) files contain information about user IP Address, User Name, URL that Referred, Access Request, Time Stamp; error codes etc. and generally resides in the web servers.

b) Apply Preprocessing:

First step is preprocessing. Preprocessing need to be carried out before performing web mining calculations on the Web server logs.

B. Preprocessing comprises of following methods [12]:

1) Data Cleansing:

During data cleansing process Irrelevant records are eliminated. Since the target of web usage mining is to retrieve traversal pattern, following two kinds of records found unnecessary and should be removed:

- 1) The records having suffixes for filename as GIF, JPEG, CSS.
- 2) While examining the web log for status field of every record, it eliminates the record with status code range more than 299 and below then 200.

2) User and Session Identification:

It's aim is to extract different user sessions from the original web access log.

- 1) If all the IP addresses are same, different users could be found by user agents and client IP address indicated by the different operation systems and browsers.
- 2) If all of the browsers, IP address, and operating systems are same, information related to referrer should be taken into account. The ReferURI is checked, new user's session is identified if the URL in the ReferURI is „-“ indicates that field hasn't been accessed before.

3) Content Retrieval:

It retrieves content derived from user's query request i.e.

cs_uri_query.forEg:Query:http://www.cs.depaul.edu/courses

/syllabus.asp?course=323-21-603&q=3&y=2002&id=671. Retrieve the content such as syllabus/courses.asp which helps in fast searching of unvisited pages i.e; pages of another user's which are similar to user's interest.

4) Path Completion:

It should be used for obtaining the complete user access path. If in the beginning of session, Referrer as well URI has some data value, eliminate value of Referrer with addition of „-“,

5) Prepare processed data base

The database which was preprocessed is now formatted and stored in the database. The data which is stored is now in the cleaned form and in well representable form.

6) *Apply Pattern mining algorithm*

Pattern mining algorithm deal with the processed data base. Processed data base is capture by the pattern mining algorithm for the generation of frequently accessed pattern.

7) *Construct Pattern tree*

Frequently accessed patterns construct a pattern tree (nodes of the tree are represented by frequently accessed patterns) for the recommendation rule generation. Sequence of pattern tree can be matched with the current access sequence of the user and then generate recommendation rule.

8) *Recommendation Rule Generation*

For the recommendation, we will apply Algorithm to the logdata to capture the list of frequently used items in the formof sets, then Calculate the Confidence and Support for different url of a website [13]. Then on the basis of the frequent data item list we will recommend (i.e. suggestion) a page to the user which is the best match of the user query.

IV. SEQUENTIAL PATTERN MINING ALGORITHMS

This algorithm was first introduced by R. Agrawal and R. Srikant [14]. To reduce the search space for finding frequent sequence they both have also proposed a generalized algorithm for sequential pattern mining GSP [15]. Closed itemset mining and Sequential pattern mining are associated with Closed sequential pattern mining. From a sequential database, total set of sequential patterns are discovered by sequential pattern mining .

A brief summary of the is given below with the help of flowdiagram.

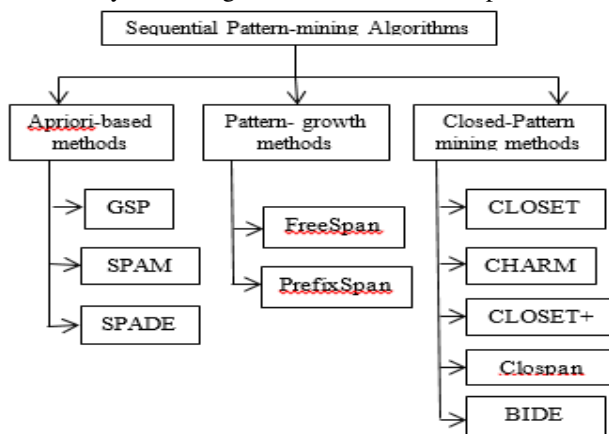


Fig. 2: Sequential Pattern-Mining Algorithms

On the basis of Apriori method algorithms such as SPAM [16], SPADE [17], and GSP were developed. Apriori-based methods use candidate generate-and-test framework, to make use of the downward closure property. Algorithms like FreeSpan [18] and PrefixSpan [19] were developed based on pattern-growth model. Algorithms like Pattern-growth follows an incremental mechanism for generating construct projections and possible frequent sequences of the database for reducing the search space.

The sequential pattern mining efficiency is increased with the help of algorithms like SPAM, SPADE and PrefixSpan, in terms of time and space complexity. SPAM and PrefixSpan work on depth-first search where as SPADE work with breadth-first search. The sequential pattern are generated using a simple join on id-lists, SPADE uses vertical data format but Prefix Span uses horizontal data format with pattern growth model. Spam has the advantage that it runs faster then SPADE and Prefix due to use of vertical bitmap representation but it also comes up with drawback of more memory consumption.

Closed itemset mining technique was developed to generate closed itemset with no supersets of same set. Closed itemset mining generates smaller set of results as compared to frequent itemset mining. Close [20] mines closed itemsets and is based on Apriori approach. At first Close employs bottom-up search and then calculates the closure of all generators.

FPclose[21] mines closed itemsets and is based on FP-growth method. During the mining process, frequency count of whole dataset store by FPtree used by FPclose. FPclose uses another data structure i.e. Closed Frequent Itemset tree by traversing all closed itemsets, to examine the closure of a frequent itemset as well as to

decrease the search space and the no of subset testing actions.

CLOSET[22] and CHARM [23] uses space efficient depth first search. To generate compressed database representation to mine closed itemset, FP tree is used by CLOSET where as CHARM uses compact vertical tid list structure also known as diffset.

CLOSET+[24] mines closed itemsets by using item skipping and subset-checking techniques. Item skipping technique speed up the mining by pruning the search space. While subset-checking technique performs the closure- checking and reduces the memory consumption. The CLOSET+ performs better from CLOSET and CHARM in terms of scalability , execution time and memory usage.

Now the recent focus in sequential pattern mining has been diverted from mining the complete set of sequential patterns to mine the closed sequential patterns for achieving a more compact output set with better efficiency. A pattern is a closed pattern if it does not have any super pattern with the same support.

For closed sequential pattern mining the two well known algorithms are CloSpan [25] and BIDE [26]. Both these techniques uses PrefixSpan framework. As the above closed itemset mining techniques, CloSpan also uses a candidate maintenance-and-test paradigm. The mining is done in two phases. In the first phase CloSpan generates a closed sequential pattern candidate set and keeps in a prefix sequence lattice where as in second phase, it eliminates non closed sequential patterns by performing post pruning. It uses an efficient search space pruning method, known as equivalence of projected databases. It performs subsequence/ supersequence checking efficiently by using the size of the projected databases as the hash key.

CloSpan produces minimum number of patterns compared to sequential pattern mining while preserving the same expressive power and runs faster than PrefixSpan. CloSpan is not scalable because it performs under candidate maintenance-and-test paradigm. In CloSpan large amount of closed sequential pattern candidates will consume more memory space utilization and it leads to large search space for checking new pattern of closure, particularly low support threshold values or long patterns.

BIDE generated closed sequential patterns without maintaining candidates. The Non closed sequences are eliminated by using BI-Directional Extension closure checking scheme. Search Space is pruned by the BackScan pruning method. It consumes less memory and runs faster compared to previously announced closed sequential pattern mining techniques, specially when the support is minimum. Scalability is better compared to previous techniques.

FMCS [27] uses breadth first search techniques and generate output on frequent closed patterns online. It removes the drawback of the candidate maintenance-and- test paradigm and generate more search space in compared to closed pattern mining algorithms explained above. It uses equivalence class to reduce the runtime.

V. RELATED WORK

As of late, to forecast navigation patterns of users and their inclination, many Web Usage Mining systems have been proposed. In this section we review some of the research work related to web recommendation.

S.Vijayalakshmi et al [28] proposed a new technique for frequent sequential pattern (FSP) mining, named as AWAPT (Adaptive Web Access Pattern Tree). To efficiently store all the sequences that contain a given item, it combines Suffix tree and Prefix tree together. During the mining, AWAPT removes recursive reconstruction of intermediate WAP tree by transmitting the binary codes to every node of it. To improve the efficiency of AWAPT, an attempt has been done to its approach. AWAPT completely removes the requirements to engage in numerous reconstructions of intermediate WAP-trees during mining and considerably decreases execution time.

Jian Pei et al [29] done survey about the problematic mining access patterns from Web logs efficiently. They have introduced a novel data structure for efficient mining of access patterns from pieces of logs, and termed as Web access pattern tree, or WAP-tree in short. it stores extremely compressed critical information for access pattern mining. And they facilitates the development of novel algorithms for mining access patterns in large set of log pieces. This algorithm can efficiently extract access patterns from web logs.

Pragya Rajput, Joy Bhattacharjee, Roopali Soni [30] proposed work helps personalize the website according to users, country being its analytical parameter (“ web personalization is a action that tailors the web experience to any individual or several users “). Distinctive IP address are provided by every internet user by the Internet Service Provider. Every region has some specific Internet service provider. For example, in India BSNL is the major ISP.

VI. PROPOSED WORK

It is a good practice if we provide a list of recommendation or suggestion to the users related to their search query, for making effective search when they interact with the web. Frequency is calculated on the basis of popularity of the searched keyword and popularity is calculated with the help of analysis of web log. In this paper we propose an

approach for the recommendation that uses pattern mining apriori algorithm for finding frequently access pattern. Apriori algorithm generates a list of frequently accessed patterns. The frequently accessed patterns are find out on the basis of the local country-wise search algorithm implemented. Then with the help of frequently accessed patterns construct a pattern tree (nodes of the tree are represented by frequently accessed patterns) for the recommendation rule generation. Diverse estimation procedures like Precision, applicability and satisfibility are considered to compute the performance of the system.

A. Sequential Pattern Mining

Generally, every line of web logs (one access record) comprises the following key information: requested URL, and HTTP status code, client IP address, date-timestamp, user ID. Prior to applying sequential pattern mining, Preprocessing must be done on the web logs. The main data preprocessing techniques used here containing session identification and user identification as well as data cleaning.

Sequential pattern mining is an significant subject of data mining which is broadly applied in several zones [31]. Generally, sequential pattern mining is defined as the determination of whole set of frequent subsequences in a set of sequences [32, 33]. Sequential pattern can be explained as the sequence of itemsets that frequently appear in a specific order, such that, all items in that same itemset are expected to have the same transaction time value or within a time gap.

In our proposed work, closed sequential pattern algorithm has been explained like this:-

1) *Algorithm 1: Call_Seq()*

1. Sequences from the database are find out.
2. Each sequence is assigned an id.
3. If either sub_sequence or super_sequence exists, then
4. Modify the pattern else add in it.
5. Scan the database.
6. If the sequence is not valid, then return.
7. Else call_seq().
8. End.

B. Pattern tree construction

A Pattern-tree model is proposed for storing sequential web access patterns compactly, so that it can be used for matching with a user's current access sequence and generating recommendation rules more efficiently in the Recommendation Rules Generation component. In general, the set of sequential web access patterns can be treated as a set of strings over a finite alphabet E (the set of unique access events). Sometimes, the edges of a Tree are labeled instead of the nodes, but nevertheless refer to the same structure in either case.

The procedure for constructing the pattern tree in the proposed system is as follows: 1) *Algorithm 2: Call_PTC()*

1. Create an empty node and set as root node.
2. For each sequence that is traversed set the current node as the root.
3. If any child node exists, then calculate the support for it and set it as the current node.
4. Otherwise, create child node with new support and set it as current node.
5. Return the created pattern tree.

C. Recommendation Rule Generation

It's elements searches for the optimum-matching access path in the Pattern tree according to the user's existing access sequence. In general, as longer the user's existing access sequence is, the lower the probability is of extracting the matching path from the Pattern-tree. Some initial items can be eliminated to make the current access sequence shorter than the depth of the Pattern-tree before the sequence matching process begins.

D. Generation of Recommendations for User Access Sequences

The constructed pattern tree is used for matching and generating web links for recommendations for user access sequences. For instance, the proposed system provides recommendations R for user web access sequences $\langle xy \rangle$ and $\langle zx \rangle$ as follows:

1) *Case 1:*

The constructed pattern tree is searched for sequential patterns with the presence of User's web access sequence

<xy>. In the pattern tree (shown in figure 3), next to the rootnode has the sequences <xyy: 2> and <xyz: 4> that are subset of given user web access sequence <xy>. So the postfixes of these sequential patterns i.e., <y> and <z> are given as recommendations to the user ($R = y: 2$ and $z: 4$).

2) *Case 2:*

For user web access sequence <zx>, the pattern tree (shown in figure 3) is searched for pattern with the presence of

<zx>. As the pattern <zx> does not have any postfixes, it's child node i.e., <y> is provided as recommendations for the user ($R = y: 2$).

3) *Algorithm 3: Call_Rec_rule_gen()*

1. Initializing the root node as null.
2. If the sequence is greater than the maximum length, then remove it from the sequence.
3. If sequence is less than minimum length, then return the root node.
4. Else set the current node as the root node.
5. If child node is available, find the child node of the current node and set the child node as current node.
6. Else remove the item from the sequence. Repeat from step 3.

The process of recommendation rule generation works in a particular algorithm that calls various algorithms within itself.

The pseudo-code for the recommendation system proposed can be explained as below:-

1. Clean the data and remove all the unnecessary data like jpeg, robots, etc.
2. Load the cleaned data.
3. Unique URLs are identified and assigning unique ids to the URLs and country names respectively.
4. Sequential mining algorithm is applied by calling Call_seq().
5. Pattern tree is constructed by calling Call_PTC().
6. User enters the prefix sequence to generate the recommendation rule by calling Call_Rec_Rule().
7. A new recommendation system is created which efficiently searches locally.

VII. PERFORMANCE EVALUATION

A. Evaluation Measures

Let $S = x_1, x_2, \dots, x_l, x_{l+1}, \dots$ can be a web access sequence. For the prefix sequence $S_{prefix} = x_1, x_2, \dots, x_l$ ($l \geq \text{MinLength}$), we generate a recommendation rule $\text{Rec_Rule} = \{s_1, s_2, \dots, s_m\}$ using the Pattern-tree, where all events are ordered by their support. If $x_{l+1} \in \text{Rec_Rule}$, Recommendation rules are labeled as correct, and incorrect. If there exists $x_i \in \text{Rec_Rule}$ ($l+1 \leq i \leq l+1+m$, $m > 0$), Recommendation rules are deemed as m-step satisfactory. Otherwise, labeled as m-step unsatisfactory. Let $R_Rule = \{\text{Rec_Rule}_1, \text{Rec_Rule}_2, \dots, \text{Rec_Rule}_n\}$ be a set of recommendation rules, where Rec_Rule_i ($1 \leq i \leq n$) is a recommendation rule. $|R_Rule| = n$ is the total number of recommendation rules in R_Rule . We define the following measures.

1) *Definition 1:*

Let R_c be the subset of R_Rule consisting of all correct recommendation rules. The overall web recommendation precision is defined as

$$\text{precision} = \frac{|R_c|}{|R_Rule|}$$

This precision measures how probable a user will access one of the recommended pages.

2) *Definition 2:*

Time Utilization = the net time utilized by the algorithm to process the code.

Where, R_Rule are total rules generated.

R_c is subset of R_Rule which are correct rules.

R_s is subset of R_Rule that consists of all satisfactory rules.



B. Results Analysis

1) Base paper results

Total number of records : 1560

CountryNo	CountryName
1	United States
2	India
3	Brazil

2) No. WebsiteName

- 1) <http://bluerobot.info/nw2/grigosikyfaheku>
- 2) <http://buttons-for-website.com>
- 3) <http://buttons-for-website.comhttp://buttons-for-website.com/>
- 4) website.com/
- 5) <http://semalt.semalt.com/crawler.php>
- 6) <http://texec.org>
- 7) <http://texec.org/>
- 8) <http://texec.org/about.html>
- 9) <https://www.megakin.com/>
- 10) <https://www.megakin.com/about.html>
- 11) <https://www.megakin.com/contact.html>

Etc.

1) Step 1:- User is asked to input the search sequence

The screenshot shows a 'Command Window' with a list of URLs and an 'Input Sequence' dialog box. The URLs listed are:

```

14 http://texec.org/services.html
15 http://whois.domaintools.com/texec.org
16 http://www.bing.com/search
17 http://www.google.co.in/url
18 http://www.google.com/m
19 http://www.google.com/search
20 http://www.google.com/url
21 http://www.gschildcare.org/
22 http://www.hackersonlineclub.com/
23 http://www.netcraft.com/survey/
24 http://www.pujyamanasmohinisandhyadevi.org/
25 http://www.texec.org/
26 http://www.texec.org/about.html
27 http://www.texec.org/contact.html
28 http://www.texec.org/css/josefin-sans/stylesheet.css
29 http://www.texec.org/portfolio.html
30 http://www.texec.org/services.html
31 https://www.google.co.in/
32 https://www.google.co.za/
33 https://www.google.com.br/
34 https://www.google.com/
35 https://www.megakin.com/
36 https://www.megakin.com/about.html
37 https://www.megakin.com/contact.html
38 https://www.megakin.com/services.html
39 https://www.pujyamanasmohinisandhyadevi.org/
40 https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheet.css
    
```

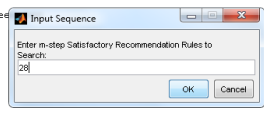
The 'Input Sequence' dialog box is open, showing the text 'Enter sequence to Search:' followed by the input '24 25 25'. There are 'OK' and 'Cancel' buttons at the bottom of the dialog box.

Elapsed time is 67.419610 seconds.

```

Command Window
27 http://www.texec.org/contact.html
28 https://www.texec.org/css/josefin-sans/stylesheets.css
29 http://www.texec.org/portfolio.html
30 http://www.texec.org/services.html
31 https://www.google.co.in/
32 https://www.google.co.za/
33 https://www.google.com.br/
34 https://www.google.com/
35 https://www.megakin.com/
36 https://www.megakin.com/about.html
37 https://www.megakin.com/contact.html
38 https://www.megakin.com/services.html
39 https://www.pujyamanasmohinisandhyadevi.org/
40 https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheets.css
Elapsed time is 67.419610 seconds.

output_rule =
    25 3
    25 3
    25 3
    25 3
    25 3
    25 3
    25 3
    25 3
    28 3
    28 3
precision = 0.750000
    
```



Elapsed time in tree creation 67.419610 seconds.

3) *Output_rule*

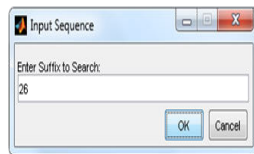
- 4) 25 3
- 5) 25 3
- 6) 25 3
- 7) 25 3
- 8) 25 3
- 9) 25 3
- 10) 28 3
- 11) 28 3

- Precision = 0.750000
Satisfaction = 0.375

2) Step 2:- User is asked to enter the suffix that he wants to search.

```

14 http://texec.org/services.html
15 http://whois.domaintools.com/texec.org
16 http://www.bing.com/search
17 http://www.google.co.in/url
18 http://www.google.com/m
19 http://www.google.com/search
20 http://www.google.com/url
21 http://www.gschildcare.org/
22 http://www.hackersonlineclub.com/
23 http://www.netcraft.com/survey/
24 http://www.pujyamanasmohinisandhyadevi.org/
25 http://www.texec.org/
26 http://www.texec.org/about.html
27 http://www.texec.org/contact.html
28 https://www.texec.org/css/josefin-sans/stylesheets.css
29 http://www.texec.org/portfolio.html
30 http://www.texec.org/services.html
31 https://www.google.co.in/
32 https://www.google.co.za/
33 https://www.google.com.br/
34 https://www.google.com/
35 https://www.megakin.com/
36 https://www.megakin.com/about.html
37 https://www.megakin.com/contact.html
38 https://www.megakin.com/services.html
39 https://www.pujyamanasmohinisandhyadevi.org/
40 https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheets.css
Elapsed time is 64.042806 seconds.
    
```



3) Step 3:- Then, user is asked to enter m-step satisfactory recommendation rules to be searched.

4) - Elapsed time in tree searching 63.0953 seconds

5) *Proposed Paper Results*

CountryNo CountryName

- 1) United States
- 2) India
- 3) Brazil



5) Country Name United States

a)WebsiteNo WebsiteName

- 1) <http://buttons-for-website.com>
- 2) <http://semalt.semalt.com/crawler.php>
- 3) <http://texec.org/>
- 4) <http://texec.org/css/josefin-sans/stylesheet.css>
- 5) <http://texec.org/services.html>
- 6) <http://whois.domaintools.com/texec.org>
- 7) <http://www.google.co.in/url>
- 8) <http://www.google.com/m>
- 9) <http://www.google.com/search>
- 10) <http://www.google.com/url>

6) Country Name India

a) WebsiteNo WebsiteName

- 1) <http://buttons-for-website.com>
- 2) <http://buttons-for-website.comhttp://buttons-for-website.com/>
- 3) <http://semalt.semalt.com/crawler.php>
- 4) <http://texec.org/>
- 5) <http://texec.org/about.html>
- 6) <http://texec.org/contact.html>
- 7) <http://texec.org/css/josefin-sans/stylesheet.css>
- 8) <http://texec.org/hoc/>
- 9) <http://texec.org/http://texec.org/>
- 10) <http://texec.org/school/>

Etc... for all the countries.

Elapsed time is 20.627226 seconds.

1) Step 1: User is asked to enter the country name first so as to search locally.

The screenshot shows a terminal window with a list of website URLs. The list includes:

```

WebsiteNo WebsiteName
1 http://buttons-for-website.com
2 http://buttons-for-website.comhttp://buttons-for-website.com/
3 http://semalt.semalt.com/crawler.php
4 http://texec.org/
5 http://texec.org/about.html
6 http://texec.org/contact.html
7 http://texec.org/css/josefin-sans/stylesheet.css
8 http://texec.org/hoc/
9 http://texec.org/http://texec.org/
10 http://texec.org/school/
11 http://texec.org/services.html
12 http://www.google.com/search
13 http://www.pujyamanasmohinisandhyadevi.org/
14 http://www.texec.org/
15 http://www.texec.org/contact.html
16 http://www.texec.org/css/josefin-sans/stylesheet.css
17 http://www.texec.org/services.html
18 https://www.google.co.in/
19 https://www.megakin.com/
20 https://www.megakin.com/about.html
21 https://www.megakin.com/contact.html
22 https://www.megakin.com/services.html
23 https://www.pujyamanasmohinisandhyadevi.org/
24 https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheet.css

```

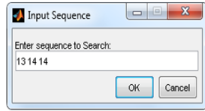
Below the terminal window, there is a dialog box titled "Input Country". It contains a text input field with "India" entered, and "OK" and "Cancel" buttons.

Country Name Brazil

2) Step 2: User is asked to input the sequence to be searched locally.

WebsiteNo	WebsiteName
1	http://buttons-for-website.com
2	http://buttons-for-website.comhttp://buttons-for-website.com/
3	http://semalt.semalt.com/crawler.php
4	http://texec.org/
5	http://texec.org/about.html
6	http://texec.org/contact.html
7	http://texec.org/css/josefin-sans/stylesheet.css
8	http://texec.org/hoc/
9	http://texec.org/http://texec.org/
10	http://texec.org/school/
11	http://texec.org/services.html
12	http://www.google.com/search
13	http://www.pujyamanasmohinisandhyadevi.org/
14	http://www.texec.org/
15	http://www.texec.org/contact.html
16	http://www.texec.org/css/josefin-sans/stylesheet.css
17	http://www.texec.org/services.html
18	https://www.google.co.in/
19	https://www.megakin.com/
20	https://www.megakin.com/about.html
21	https://www.megakin.com/contact.html
22	https://www.megakin.com/services.html
23	https://www.pujyamanasmohinisandhyadevi.org/
24	https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheet.css

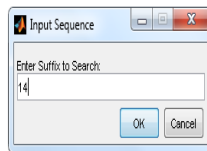
Country Name	Brazil
--------------	--------



3) Step 3: Enter the suffix for the local search.

WebsiteNo	WebsiteName
1	http://buttons-for-website.com
2	http://buttons-for-website.comhttp://buttons-for-website.com/
3	http://semalt.semalt.com/crawler.php
4	http://texec.org/
5	http://texec.org/about.html
6	http://texec.org/contact.html
7	http://texec.org/css/josefin-sans/stylesheet.css
8	http://texec.org/hoc/
9	http://texec.org/http://texec.org/
10	http://texec.org/school/
11	http://texec.org/services.html
12	http://www.google.com/search
13	http://www.pujyamanasmohinisandhyadevi.org/
14	http://www.texec.org/
15	http://www.texec.org/contact.html
16	http://www.texec.org/css/josefin-sans/stylesheet.css
17	http://www.texec.org/services.html
18	https://www.google.co.in/
19	https://www.megakin.com/
20	https://www.megakin.com/about.html
21	https://www.megakin.com/contact.html
22	https://www.megakin.com/services.html
23	https://www.pujyamanasmohinisandhyadevi.org/
24	https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheet.css

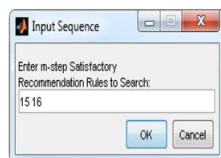
Country Name	Brazil
--------------	--------



4) Step 4:- Then, user is asked to enter m-step satisfactory recommendation rules to be searched.

WebsiteNo	WebsiteName
1	http://buttons-for-website.com
2	http://buttons-for-website.comhttp://buttons-for-website.com/
3	http://semalt.semalt.com/crawler.php
4	http://texec.org/
5	http://texec.org/about.html
6	http://texec.org/contact.html
7	http://texec.org/css/josefin-sans/stylesheet.css
8	http://texec.org/hoc/
9	http://texec.org/http://texec.org/
10	http://texec.org/school/
11	http://texec.org/services.html
12	http://www.google.com/search
13	http://www.pujyamanasmohinisandhyadevi.org/
14	http://www.texec.org/
15	http://www.texec.org/contact.html
16	http://www.texec.org/css/josefin-sans/stylesheet.css
17	http://www.texec.org/services.html
18	https://www.google.co.in/
19	https://www.megakin.com/
20	https://www.megakin.com/about.html
21	https://www.megakin.com/contact.html
22	https://www.megakin.com/services.html
23	https://www.pujyamanasmohinisandhyadevi.org/
24	https://www.pujyamanasmohinisandhyadevi.org/css/josefin-sans/stylesheet.css

Country Name	Brazil
--------------	--------



7) Output_rule



14 3
 14 3
 14 3
 14 3
 14 3
 14 3
 14 3
 16 3
 16 3

Precision = 0.625000

Satisfaction = 0.125

Elapsed time in tree searching is 10.894232 seconds.

VIII. CONCLUSION

In this paper we have discussed about the sequential pattern mining and web mining. Various sequential mining algorithms have been developed that have been developed till date with all the pros and cons of every algorithm. Recommender system with various steps used in this system has also been discussed. A brief introduction to each algorithm has been discussed.

REFERENCES

- [1] U. M. Fayyad, G. P. Shapiro and P. Smyth, From Data Mining to Knowledge Discovery in Databases. 0738-4602-1996, AI Magazine (Fall 1996). pp: 37–53.
- [2] J. Han and M. Kamber, Data Mining: Concepts and Techniques. Second edition Morgan Kaufmann Publishers.
- [3] S.K. Pani, L.Panigrahy, V.H.Sankar, Bikram Keshari Ratha, A.K.Mandal, S.K.Padhi, “Web Usage Mining: A Survey on Pattern Extraction from Web Logs”, International Journal of Instrumentation, Control & Automation
- [4] M. Eirinaki and M. Vazirgiannis, “Web mining for web personalization”, ACM Transactions on Internet Technology, Vol. 3, No. 1, 2003, pp. 1-27.
- [5] Dunham, M. H. 2003. Data Mining Introductory and Advanced Topics. Pearson Education.
- [6] Ms.Dipa Dixit and Mr Jayant Gadge, “Automatic Recommendation for Online Users Using Web Usage Mining”, International Journal of Managing Information Technology (IJMIT) Vol.2, No.3, August 2010.
- [7] Burke, R.: Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12(4), 331–370 (2002)
- [8] Linden, G., Smith, B., York, J.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 7(1), 76–80 (2003)
- [9] Balabanovic, M., Shoham, Y.: Fab: Content-based, Collaborative Recommendation. Communications of the ACM 40(3), 66–72 (1997)
- [10] Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: The Adaptive Web, pp. 291–324. Springer Berlin / Heidelberg (2007)
- [11] Schafer, J.B., Konstan, J.A., Riedl, J.: E-commerce recommendation applications. Data Mining and Knowledge Discovery 5(1/2), 115–153 (2001)
- [12] M. Perkowitz and O. Etzioni, "Towards adaptive Web sites: Conceptual framework and case study," Artificial Intelligence, vol. 118, pp. 245-275, 2000.
- [13] M. Jalali, N. Mustapha, A. Mamat, Md N. Sulaiman, “OPWUMP An architecture for online predicting in WUM-based personalization system”, In 13th International CSI Computer Science, Springer Verlag, 2008. 307.
- [14] R.Agrawal and R.Srikant, “Mining Sequential Patterns,” In Proceedings of International conference on data engineering. pp. 3-14.
- [15] Minghua Zhang, Ben Kao, Chi-Lap Yip, David Cheung, “A GSP-based Efficient Algorithm for Mining Frequent Sequences”
- [16] J. Ayres, J. Gehrke, T. Yiu, and J. Flannick, “Sequential pattern mining using a bitmap representation,” Proceedings of ACM SIGKDD ‘02, pp. 429-435, July 2002.
- [17] M. Zaki, “SPADE: An efficient algorithm for mining frequent sequences,” Kluwer Academic Publisher’s Machine Learning, vol. 42, pp. 31-60, 2001.
- [18] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.C. Hsu, “FreeSpan: Frequent pattern projected sequential pattern mining,” Proc. ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (SIGKDD ‘00), pp. 355-359, Aug. 2000.

- [19] J. Pei, J. Han, B. Mortazavi-Asl, and et al., "PrefixSpan : Mining sequential patterns efficiently by prefix-projected pattern growth," Proc. Int'l Conf. Data Engineering (ICDE '01), pp. 215-224, Apr. 2001.
- [20] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Discovering frequent closed itemsets for association rules," 7th Int'l Conf. on Database Theory, pp. 398-416, Jan. 1999.
- [21] Grahne G and Zhu J F, "Fast algorithm for frequent itemset mining using FP-Trees," IEEE Trans. on Knowledge and Data Engineering, vol. 17, no. 10, pp. 1347-1362, 2005.
- [22] J. Pei, J. Han, and R. Mao, "CLOSET: An efficient algorithm for mining frequent closed itemsets," Proceedings of ACM DMKD '00, pp. 21-30, May 2000.
- [23] M. Zaki and C. Hsiao, "CHARM: An efficient algorithm for closed itemset mining," Proceedings of SIAM's SDM '02, pp. 457-473, Apr. 2002.
- [24] J. Wang, J. Han, and J. Pei, "CLOSET+: Searching for the best strategies for mining frequent closed itemsets," Proceedings of ACM SIGKDD '03, pp. 236-245, Aug. 2003.
- [25] X. Yan, J. Han, and R. Afshar, "CloSpan: Mining closed sequential patterns in large databases," Proceedings of SIAM's SDM '03, pp. 166-177, May 2003.
- [26] J. Wang, J. Han, and Chun Li, "Frequent closed sequence mining without candidate maintenance," IEEE TKDE, vol. 19, no. 8, pp. 1042-1056, Aug. 2007.
- [27] Nancy P. Lin, Wei-Hua Hao, Hung-Jen Chen, Hao-En Chueh and Chung-I Chang, "Fast mining of closed sequential patterns," WSEAS Transactions on Computers, vol. 7, no. 3, Mar. 2008.
- [28] S.Vijayalakshmi, V.Mohan and S.Suresh Raja, "Mining Of Users" Access Behaviour For Frequent Sequential Pattern From Web Logs", International Journal of Database Management Systems (IJDMS) Vol.2, No.3, August 2010.
- [29] Jian Pei, Jiawei Han, Behzad Mortazavi-asl, and Hua Zhu, "Mining Access Patterns Efficiently from Web Logs", Natural Sciences and Engineering Research Council of Canada (grant NSERC-A3723), the Networks of Centres of Excellence of Canada (grant NCE/IRIS-3), and the Hewlett-Packard Lab.
- [30] Pragya Rajput, Joy Bhattacharjee, Roopali Soni, "A Proposed Framework to Implicit Measures of User Interests through Country and Predicting Users" Future Requests in WWW", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013.
- [31] Sizu Hou, Xianfei Zhang, "Alarms Association Rules Based on Sequential Pattern Mining Algorithm," In proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 2, pp.556-560, Shandong, 2008.
- [32] R. Agrawal and R. Srikant, "Mining Sequential Patterns", In Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, Taipei, Taiwan, 1995.



Impact Factor: 8.423



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details