International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Predicting Users Behavior from the Analysis of Web Server logs Using Hash Map and PAFI Algorithm

Ms.C.Thavamani[1], Dr.A.Rengarajan[2]
[1] Research scholar, Bharathiar University, Coimbatore, Tamil Nadu, India
[e-mail:thava.chenni@gmail.com]
[2] Professor, Department of CSE, Veltech Multitech Dr.RS Engineering College, Avadi, Chennai, Tamil Nadu, India
[e-mail: rengu_rajan@yahoo.com]

ABSTRACT- A Weblogs contains series of transactions updated frequently, while users accessing the websites. It comprises of various entries like IP address, status code and number of bytes transferred, categories and time stamp. The user interest can be classified based on categories and attributes and it is helpful in identifying user behavior. The log query parser is to convert unstructured log to structured log based on user interest. The weblog data can be classified as successful and unsuccessful data. The aim of the research is to classify the data of success response and analyze the user navigation. The process of identifying user behavior consisting of data collection, query parser, pre-processing and pattern analysis that will help us to analyze and predict the user behavior in short time. This research work explores to analyze the user prediction, based on the user preference present in various levels that is captured from weblogs.

Keywords: User navigation, web mining, user behavior, traversal pattern, prediction accuracy, Data Mining.

Introduction

In current situation explosive growth of knowledge available on internet makes the users to access the information day by day. It becomes much more difficult for users to access relevant information efficiently. Analyzing and modeling web navigation behavior is helpful in understanding demands for online users. Web mining is a data mining technique to extract and analyze useful information from web data [5]. Based on the kind of data, web mining can be classified in to three different categories namely web content mining, web usage mining and web structure mining. Web content mining is the discovery of contents from web documents such as image, text, audio, video etc. Web structure mining focus on analyzing the physical link structure of websites. Web usage mining analyzes the browsing activity.

The web usage data consists of data from weblog. The user accessing information from the websites is stored as logs. The log contains series of user transactions which are frequently updated whenever the user accesses the website [6]. The

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

prediction of user behavior can be identified only through logs. The weblog contains unstructured format, so convert to raw weblog to processed weblog using data preprocessing [7], the data preprocessing includes Data Cleaning, User identification, Session Identification, content retrieval and path completion to get user navigation pattern.

Every Manufacturing industry is interested in predicting vehicle preferences by customers. Most of the transactions are lie on website and then registered website use the customer prediction. The Vehicle prediction is analyzed by using the attributes and categories in weblog, from those different vehicles can be classified based on user preference. The prediction process is consists of online and offline with respect to web server activity [8]. Offline data analyze historical data, such as log file or weblogs which are captured from server. Weblog used in online, with the help of user intuition list, whenever that user comes online for next time.

The process of identifying user behavior is based on previous user navigation that is taken from weblogs; through weblog web traversal pattern is discovered. The classifier classifies the WTP into frequent sequence and semi-frequent sequence from that user buying behavior is analyzed.

Related Work

Web usage mining predict the user navigation behavior based on the preferences in website, User navigation technique uses data preprocessing Dr. A. R. Patel and Renata Ivancsy [5] suggested that the weblog contains raw log format, so convert to unprocessed weblog to processed weblog using data preprocessing, the data preprocessing technique contains Data Cleaning, User identification, Session Identification, content retrival and path completion to get user navigation pattern. After getting the processed log, the given log is converted in to sequence of patterns based on user's pattern.

Hong Cheng and Xifeng Yan [6] showed user get the processed log from raw log, then web traversal pattern is analyze to classification algorithm to classify the processed log into maximum forward sequence. The algorithms classify the Frequent Sequence, Semi-frequent Sequence and In-frequent Sequence. D. Kerana Hanirex and Dr. M. A. Dorai Rangaswamy [8] showed classification of frequent item sets, semi-frequent item sets and In-frequent item sets and database transactions into clusters. Clusters are grouped based on the similar traversal pattern. Then it finds the frequent item sets, semi-frequent item sets and In-frequent item sets, the clusters reduces the number of transaction in the database and the efficiency is improved.

Mobasher[15] suggested Web personalizer is a dynamic recommendation from a list of hypertext links to users. The web personalizer shows the usage of data and also shows the hyperlinks of the website. Weblogs are classifying in to frequent, semi-frequent and infrequent sequences, based on the classification the sequences are converted to form clusters. The online phase considers the maximum active user session in order to find matches of most user activities and discovered

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

usage of user profiles. A web usage mining suggest to provide useful information and the web user navigation technique is to optimize the server performance based on data preprocessing is a technique to cleaning, user identification, session identification, classification and clustering that will helps to understand the user's behavior.

Joachims, Juhne [29] and Pazzani and Billsus [31] showed the examples of visitors suggest links of each user. The user visit the web server based on user interest, by the next time the browser will show the user suggest website based on user interest. The browser will follow the frequent viewed items, whenever the user will visit some other pages that will show frequently of webserver links, then only user can follow the frequently visited links.

Fu and Perkowitz and Etzioni [32] suggested that the user will visit the web server by current and past sessions. The algorithm suggest that first top m pages based on user's most visited current sessions, the system shoud analyse the maxpath and minpath of navigated sessions, then only the visitor should easily navigate the path from the list of max list of path, by selecting min path, the algorithm select the best path from the list of frequent path.

The user prediction can be analyzed based on the previous user navigation path, to find the navigated path. The following algorithm must be analyzed such as Query log parser, user identification, session identification, pattern generation, and pattern analysis, by the following technique user prediction will be analyzed. Cadez, 2000 [33] suggested markov model is to analyze the previous results to find the user behavior, the model analyze various results based on previous user path navigated, it reduces the number of paths and reduces the clustering size.

Framework for User Navigation System

Figure 1, shows the framework for analysis of user behavior pattern in web. It consists of five phases

- Acquisition of weblogs
- Log Query Parser
- Preprocessing
- Navigation Pattern Modeling
- Clustering
- Classifier

Acquisition of Weblogs

While the user accesses the web server, then the transactions are recorded in the form of weblogs. The web log consists of history of user navigation that is stored in web server and the log consists of following attributes

1. the user's IP address
2. the remote log name of the user

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

3. the access date and time
4. the request method
5. the URL of the page
6. the protocol (HTTP 1. 0, HTTP 1. 1, etc.)
7. the return code
8. categories
9. the number of bytes transmitted

The above attributes helps to identify the user behavior and this can be classified after preprocessing.

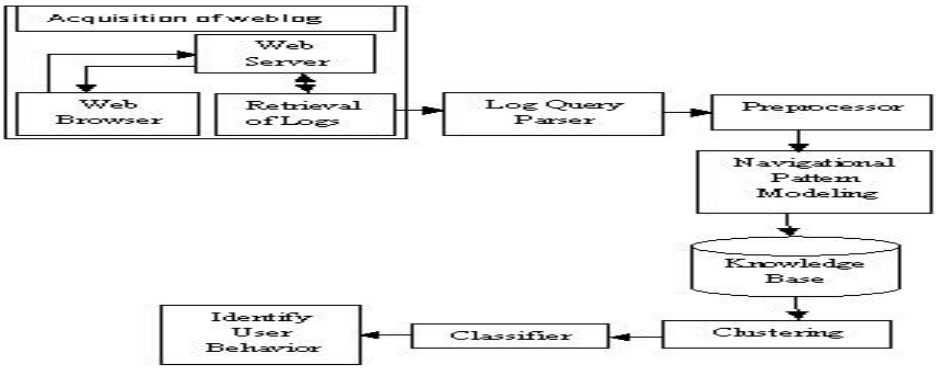
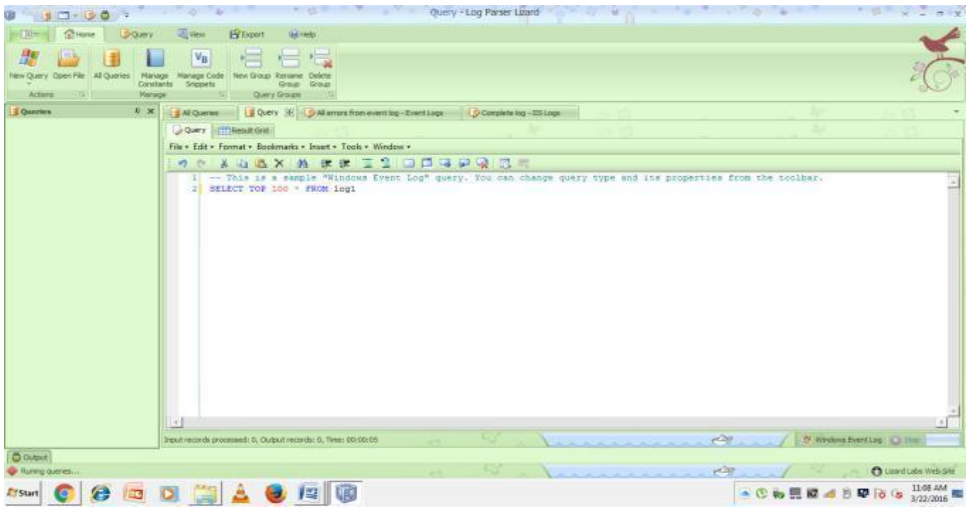


Figure. 1. Framework for Analyzing User Behavior

Log Query Parser

The log query parser is to extract unstructured log to structured log based on user interest. This parser provides universal query access to text based data such as log files, XML files and CSV files.



International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Figure 1. log query parser

Preprocessing

The log contains unstructured format of user navigation information, so conversion of unstructured log to structured log is required. It can be done through data preprocessing technique, which consists of three phases

- Data cleaning
- User identification
- Session identification

1	199.72.81.55 - - [01/Jul/1995:00:00:01 -0400] "GET /history/apollo/ HTTP/1.0" 200 6245
2	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
3	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
4	205.212.115.106 - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
5	129.94.144.152 - - [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
6	129.94.144.152 - - [01/Jul/1995:00:00:17 -0400] "GET /images/ksclogo-medium.gif HTTP/1.0" 304 0
7	199.120.110.21 - - [01/Jul/1995:00:00:17 -0400] "GET /images/launch-logo.gif HTTP/1.0" 200 1713
8	205.189.154.54 - - [01/Jul/1995:00:00:24 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
9	205.189.154.54 - - [01/Jul/1995:00:00:29 -0400] "GET /shuttle/countdown/count.gif HTTP/1.0" 200 40310
10	205.189.154.54 - - [01/Jul/1995:00:00:40 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 200 786
11	205.189.154.54 - - [01/Jul/1995:00:00:41 -0400] "GET /images/KSC-logosmall.gif HTTP/1.0" 200 1204

Figure. 2. Acquisition of weblog

Data Cleaning

Let us consider the access log of NASA web site. The users were allowed to browse their own interest. These logs have been recorded for the period of 01/Jul/1995;00:00:01 to 14/Jul/1995;00:00:01, logs details were acquired and preprocessed for further navigation prediction. The log file contains 2486 records, in that each record having the status code. The status code decides the success and failure of webpage. The status code between 200 to 400 is valid. The data cleaning process removes the records with graphics and video format such as gif, jpeg etc., which is not helpful for identifying user navigation. Hence valid records are considered for further identification. After the cleaning process in log 846 records are obtained. For navigation mapping of pages with identify is mapped semantically. From the above log P1 consider as Home page, P2 as Missions, P3 as Galleries, P4 as NASA TV, P5 as Follow NASA, P6 as Downloads, P7 as About, P8 as NASA Audiences, P9 as Topics etc.

1) Exclusive User Identification

2)

Traditional user identification is carried out according to some specific rules. Different IP address refers to different users. The same IP with different operating system or different browser should be considering as different user. While the IP, operating system and browsers are all the same, new user can be determined whether the requesting page can be reached by accessed pages before according to the topology of the site.

Exclusive user identification is important process next to data cleaning and sessionization. Distinct users are identified based on the rules suggested in User Identification section. Though many efficient algorithms are there, those are fail in

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

accuracy and efficiency when the size of the Log Database increases. Today's modern web servers are capable of handling terabytes of data conventional algorithms are outmoded in handling these scenarios. Considering the above facts this study has proposed an efficient Exclusive User Identification algorithm that uses modern Hashing techniques to identify unique user quickly inspire the huge size of the database. A new hashing key has been proposed and successfully implemented in the algorithm to locate the user.

Techniques and Usage of Hash Map

For a huge database structure, it can be almost next to impossible to search all the index values through all its level and then reach the destination data block to retrieve the desired data. Hashing is an effective technique to calculate the direct location of a data record on the disk without using index structure. Hashing uses hash functions with search keys as parameters to generate the address of a data record.

Hash Map Organization

Hash Map is organized its data with 'n' buckets to store all records in a Log based on its predefined hash function. This helps to locate the specific user IP address more faster.

- **Bucket** − A hash file stores data in bucket format. Bucket is considered a unit of storage. A bucket typically stores one complete disk block, which in turn can store one or more records.
- **Hash Function** − A hash function, **h,** is a mapping function that maps all the set of search-keys **K** to the address where actual records are placed. It is a function from search keys to bucket addresses.

Dynamic Hashing

The problem with static hashing is that it does not expand or shrink dynamically as the size of the database grows or shrinks. Dynamic hashing provides a mechanism in which data buckets are added and removed dynamically and on-demand. Dynamic hashing is also known as **extended hashing** . Hash function, in dynamic hashing, is made to produce a large number of values and only a few are used initially.

Considering this actuality, we presented a new algorithm called "EUI (EXCLUSIVE USER IDENTIFICATION)". It analyses more factors, such as user's IP address, Web site's topology, browser's edition, operating system and referrer page. This algorithm possesses preferable precision and expansibility. Proposed method shows comparison not only based on User_IP. Sometimes same User_IP may generate the different web users, based on path which chosen by any user and access time with referrer page we find out the distinct web user. When huge databases were taken for consideration the time taken to locate the records was much, hence appropriate methodology has to be incorporated to make the process faster. Taking these prevailing conditions, the study has proposed a new Hashing formulation, to minimize the searching time for large datasets. The formation of the Hashing technique is discussed below.

Proposed Hash Map Function is **N mod 2 * K**

Where N refers the record number indirectly pointing the data (an IP address or an Operating system or a browser) and K refers to the virtual address of the bucket. The multiplied factor gives the original location of the data. Substitute $Nmod_2$ with parameter H the above equation becomes **H(K).**

Proposed Algorithm for Exclusive User Identification is as follows,

Definition: given a clean and filtered web log file and record set web log file
Records RS= {r1,r2,r3……rn} where n>0

Step1: Input Log database RUser of N records
Step2: Exclusive User identification base
Step3: RUser=P<url, ip_addr, agent, method, operating system,status,session id,time_stamp>
Step4: RS=<r1,r2,r3…rn> where n!=0,i=0
Step5: While (i<n)
Step 6: While (Log database<>eof)

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Step7: Read Log database RS
Step8 : Substitute the proposed Hash Function **RS(i)mod$_2$ * K(i)**. If userip or agent not part of Exclusive user identification base then It is treated as new user and copy userip and agent in Exclusive user Identification base.
Step9: End if
Step 10 : End loop(Log database)
Step11: i=i+1;
Step12: End loop (Web log file)
Setp13: End

Session Identification

After the data cleaning and user identification, the navigation pattern mining classifies the user based on session. The important operation of navigation pattern mining is to cluster the session.

Navigation pattern Modeling

After the preprocessing of web server log file, data mining techniques are applied. The sequence of pattern is improved from preprocessor technique, it contains forward reference. The sub sequences can be generated by maximum forward algorithm; it contains both forward and backward reference. The directed graph in which N nodes represent N web pages

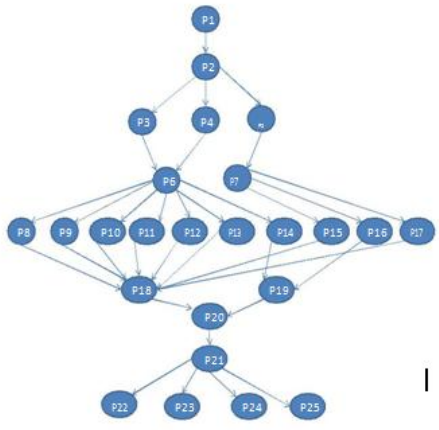


Figure 3. A Decision Tree Generation for Interested User and Not Interested User

The Decision tree is generated from the weblogs of NASA website. The pattern classifies the interested users and not interested users. The patterns supports organization in decision making process.

Table 1 the browsing pattern of user behavior

Pattern Num	Browsing Pattern
1	< P1, P2, P3, P6, P8, P18, P20, P21, P22>
2	<P1, P2, P3, P6, P10, P18, P20, P21, P23>
3	<P1, P2, P5, P7, P17, P7, P16>
4	<P1, P2, P4, P7, P16, P19, P20, P21, P25>
5	<P1, P2, P4, P7, P15, P19, P20, P21, P24>
6	<P1, P2, P3, P7, P11, P6, P9>
7	<P1, P2, P4, P6, P4, P7>

Clustering

The Partition Algorithm Frequent itemset (PAFI) is a clustering technique to cluster the browsing pattern of users. The table 1 has been generated based on decision tree; the table shows the sequence of user navigation pattern that is taken from weblogs. The each pattern is considered as one session.

The PAFI algorithm clusters the similar sequence that is taken from browsing pattern. The table 2 shows the pattern and count of browsing pattern that is helpful to identify the user behavior.

Table 2 Count of browsing pattern

Pattern Id	Pattern	Count
1	< P1, P2, P3, P6, P8, P18, P20, P21, P22>	16
2	<P1, P2, P4, P7, P15, P19, P20, P21, P24>	17
3	<P1, P2, P4, P7, P16, P19, P20, P21, P25>	11
4	<P1, P2, P3, P6, P10, P18, P20, P21, P23>	16
5	<P1, P2, P5, P7, P17, P7, P16>	2

Step1: Begin Number of clusters (NOC) = count of transactions (COT)/N //N is random natural number
Step 2: FOR i= 1 to NOC DO BEGIN

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Step 3: FOR each cluster Ci DO
BEGIN
Step 4: FOR each transaction tÎD DO BEGIN
Step 5: Find t such that t having highest number of items Step 6: Put t in Ci
Step 7: END
Step 8: END
Step 9: Return Clusters with 1 item set.

Classifier

The Classifier utilizes browsing pattern from table 2. Using IncSpan algorithm, the browsing pattern is converted in to frequent sequence, semi frequent sequence and in frequent sequence.

The user behavior is classified as interested users and not interested users. The interested user can be classified based on the count of pattern. If the count of the pattern is greater than 10, then the sequence is frequent sequence, so the user is interested in purchasing the vehicle. If the count is in-between 5 to 10 then the sequence is semi frequent, otherwise the sequence is in frequent, so the user is not interested in purchasing.

Table 3 Interested and Not Interested Users prediction

Pattern Id	Pattern
1	< P1, P2, P3, P6, P8, P18, P20, P21, P22>
2	<P1, P2, P4, P7, P15, P19, P20, P21, P24>
3	<P1, P2, P4, P7, P16, P19, P20, P21, P25>
4	<P1, P2, P3, P6, P10, P18, P20, P21, P23>
5	<P1, P2, P5, P7, P17, P7, P16>

The table 3 shows the prediction of interested and not interested users of various concepts of space.

D-Original database
D'-Appended database min_sup-threshold µ-Buffer ratio

FS-set of frequent sequences SFS-set of semi-frequent sequences
>min_sup-sequence is frequent
µ <= 1-sequence is semi-frequent
< µ * min_sup-sequence is in frequent

Input: An appended database D', min_sup, µ, frequent sequences FS in D, semi-frequent sequences SFS in D.
Output: FS' and SFS'
Step 1: FS' = Ø, SFS' = Ø
Step 2: Scan LDB for single items;
Step 3: Add new frequent item into FS';
Step 4: Add new semi-frequent item into SFS'; Step 5: for each new item I in FS' do
Step 6: PrefixSpan(I, D'|i, µ * min_sup, FS', SFS'); Step 7: for every pattern p in FS or SFS do
Step 8: check sup(p);
Step 9: if sup(p) = sup(p) + sup(p) >= min_sup
Step 10: insert(FS', p);
Step 11: if sup(p) >= (1-µ)min_sup
Step 12: PrefixSpan(p, D'|p, µ * min_sup, FS', SFS'); Step 13: Else
Step 14: insert (SFS', p); Step 15: return;

Conclusion

The proposed architecture is implemented by using Windows XP, SQL Server 2000 and R language 3.3.2 tools and PAFI algorithm, the system will predict the user's next access page based on user interest and improves prediction accuracy, it utilizes the previous user navigation results and classify the frequent item set, semi-frequent item set and in-frequent item set and reduces the number of path from the weblogs. The system will concentrate to extend with Fuzzy algorithm to identify the buying prediction behavior.

References

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

[1] Md. Ezaz Ahmed, Dr. Y. K. Mathur and Dr. Varun umar, "Knowledge Discovery in Health Care DatasetsUsing Data Mining Tools", International Journal of Advanced Computer Science and Applications, vol. 3, No. 4, 2012.

[2] Heena Akbar, Debra Anderson and Danielle Gallego, "Predicting intentions and behavior in Populations with or at-risk of diabetes:a systematic review", http://creativecommon. org/licenses/by-nc-nd/4. 0/), 2015.

[3] Kerstin Denecke and Wolfgang Nejdl, "How valuable is medical data? Content analysis of the medical web", Information Science 2009, Elsevier.

[4] M. Perkowitz and O. Etzioni. Adap-tive websites: an AI challenge. In Proc. 15th Intl. Joint Conf. on Art. Int., 1997.

[5] Ketul B. Patel, Dr. A. R. Patel, "Process of web usage mining to find interesting patterns from web usage data, "www. ijctonline. com vol. 3, no. 1, Aug 2012

[6] Hong Cheng, Xifeng Yan, Jiawei HanIncSpan: Incremental Mining of Sequential Patterns in large database.

[7] Pawel Weichbroth, Mieczyslaw Owoc, Michal Pleszkun "web user navigation patterns discovery from WWW server Log Files", IEEE 2012.

[8] D. Kerana Hanirex, Dr. M. A. Dorai Rangaswamy International Journal on Computer Science and Engineering, "Efficient Algorithm for Mining Frequent Itemsets Using Clustering Technique, "vol. 3, no. 3, March 2011.

[9] X. Fang and C. Holsapple, "An Empirical Study of WebSite Navigation Structures' Impacts on Web Site Usability, "Decision Support Systems, vol. 43, no. 2, pp. 476-491, 2007.

[10] J. Lazar, Web Usability: A User-Centered Design Approach. Addison Wesley, 2006.

[11] D. F. Galletta, R. Henry, S. McCoy, and P. Polak, "When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth, " Information Systems Research, vol. 17, no. 1, pp. 20-37, 2006.

[12] J. Palmer, "Web Site Usability, Design, and Performance Metrics, "Information Systems Research, vol. 13, no. 2, pp. 151-167, 2002.

[13] V. McKinney, K. Yoon, and F. Zahedi, "The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach, " Information Systems Research, vol. 13, no. 3, pp. 296-315, 2002.

[14] T. Nakayama, H. Kato, and Y. Yamane, "Discovering the Gapbetween Web Site Designers' Expectations and Users' Behavior, "Computer Networks, vol. 33, pp. 811-822, 2000.

[15] M. Perkowitz and O. Etzioni, "Towards Adaptive WebSites:Conceptual Framework and Case Study, " Artificial Intelligence, vol. 118, pp. 245-275, 2000.

[16] J. Lazar, User-Centered Web Development. Jones and Bartlett Publishers, 2001.

[17] Y. Yang, Y. Cao, Z. Nie, J. Zhou, and J. Wen, "Closing the Loop in Webpage Understanding, " IEEE Trans. Knowledge and Data Eng., vol. 22, no. 5, pp. 639-650, May 2010.

[18] J. Hou and Y. Zhang, "Effectively Finding Relevant Web Pagesfrom Linkage Information, " IEEE Trans. Knowledge and Data Eng., vol. 15, no. 4, pp. 940-951, July/Aug. 2003.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 6, 2017 ISSN: 2348-6600
Reference ID: IJCS-182 PAGE NO: 1104-1113
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

[19] H. Kao, J. Ho, and M. Chen, "WISDOM:Web IntrapageInformative Structure Mining Based on Document Object Model, "IEEE Trans. Knowledge and Data Eng., vol. 17, no. 5, pp. 614-627, May 2005.

[20] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on Web usage mining"Communications of the ACM, vol. 43, pp. 142-151, 2000.

[21] C. R. Anderson, P. Domingos, and D. S. Weld, "Adaptive Web Navigation for Wireless Device" Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, pp. 879-884, 2001.

[22] I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White, "Visualization of navigation patterns on a Web site using model-based clustering" Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 280-284, 2000.

[23] Dr. R. Lakshmipathy, V. Mohanraj, J. Senthilkumar, Y. Suresh, "Capturing Intuition of Online Users using a Web Usage Mining" Proceedings of 2009 IEEE International Advance Computing Conference (IACC 2009)Patiala, India, 6-7 March 2009.

[24] J. Ben Schafer, Joseph Konstan, John Riedl "Recommender Systems in E-Commerce" GroupLens Research Project Department of Computer Science and Engineering, University of Minnesota.

[25] 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

[26] R. Baraglia and F. Silvestri, "Dynamic personalization of web sites without user intervention, " Communications of the ACM, vol. 50, pp. 63-67, 2007.

[27] I. V. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White. Visualization of navigation patterns on a web site using model based clustering. In Proc. 6th Intl. Conf. on Knowl-edge Discovery and Data Mining, 2000.

[28] R. R. Sarukkai. Link prediction and path analysis using Markov chains. In Proc. 9th Intl. WWW Conf., 2000.

[29] T. Joachims, D. Freitag, and T. Mitchell. WebWatcher: A tour guide for the World Wide Web. In Proc. 15th Intl Joint Conf. on Art. Int., 1997.

[30] J. Juhne, ¨ A. T. Jensen, and K. Grønbæk. Ari-adne: a Java-based guided tour system for the World Wide Web. In Proc. 7th Intl. WWW Conf., 1998.

[31] M. J. Pazzani and D. Billsus. Adaptive web site agentsIn Proc. 3rd Intl. Conf. on Autonomous Agents, 1999.

[32] X. Fu, J. Budzik, and K. J. Hammond. Mining navigation history for recommendation. In Proc. 2000 Conf. on Intelligent User Interfaces, 2000.

[33] M. Perkowitz and O. Etzioni. Adap-tive websites: an AI challenge. In Proc. 15th Intl. Joint Conf. on Art. Int., 1997.