Web Personalization Based on Enhanced Web Access Pattern using Sequential Pattern Mining

Kuldeep Singh Rathore[1] Sanjiv Sharma[2]

[1]Research Scholar, CSE/IT Department, MITS Gwalior, India
[1]Email id - kuldeeprathore22@gmail.com
[2]Assistant Prof., CSE/IT Department, MITS Gwalior, India
[2]Email id - er.sanjiv@gmail.com

Abstract: *The development of the web has created a big challenge for directing the client to the website pages according to their need. Accordingly, just option is to capture the intuition of the client and provide them a list of recommendation. More specifically, the behavior of an online navigation develops with each passing day; consequently extracting information intelligently from it is a tedious task. Webmaster of an organization ought to utilize methods of web mining to fetch intuition, WUM is one among them. Web usage mining is planned to operate on web server logs; logs contain client's behavior towards browsing patternon web which is very useful for the web recommendation. Recommendation is an application of WUM. Consequently, recommendation system can be utilized to forecast the navigation pattern of client and recommend those to client in a recommendation list form. In this paper, we tend to propose a recommendation approach that recommends a listing of pages based mostly upon client's historic pattern (recorded within the web log). This approach brings the advance within the accuracy of the pages displayed to the client or users.*

Keywords: Web mining, Web usage mining, Sequential pattern mining, Recommender system, Rule generation.

1. INTRODUCTION

There is a huge influence of method on our present life. Electronic devices, computers and multimedia are things and deal with everyday. Particularly Internet is becoming most significant for closely everyone as it is one of the newest and most forward-looking media and confidently "the" medium of the future. Therefore we assumed that it would be essential to the think about few good and bad features of how this medium influences us, what effects it has on our behaviour social and what the future will look like. The Internet has changed our life extremely, there is no doubt about that. There are numerous Internet benefits that present you the new medium significance.

Web mining is the data mining methods application to the Web data. Web mining is quite useful to resolve the problem of finding out that how users are using web sites. It involves mining or analyzing logs and the steps that have to be follow to get significant data from web logs data collection, pre-processing, data enrichment and pattern analysis and discovery. Data mining techniques application to the WWW, referred to as web mining, has been focus of various present research projects and various papers. However, there is no recognized vocabulary, leading to confusion when comparing research efforts. The term Web mining has been used in two different distinct ways. The first one is as web content mining, the information discovery process from sources across the WWW. The second one is known as web usage mining, is the mining procedure for access patterns and user browsing. In this paper we describe Web mining and present summary of the numerous research problems, methods, and development efforts. The internet is flooded with a huge valuable and useless knowledge. It is most hard to describe valuable knowledge for a specific user which is variable from time to time. The valuable knowledge of one specific time may not be valuable on various time or a several condition. The web itself is concerning currently with newer technologies. Since internet is using freewares that receives structured, non-structured, ordered, non-ordered format to represent knowledge in the web, finding not only the relevant data but to plan them according to the user interest is also a key challenge presently and is called as web personalization.

Web personalization is the tailoring methodology a website or content [1] of website to the necessities of all specific user or users set, taking data advantage attained by the exploration of the customer's navigational performance [2]. Personalization means to provide facilities to all particular user in a tailored manner. The propagation of information on the internet has ended the personalization system an obligation. The personalization technique must have the capability to resolve the additional data problems and let the clienteles rehearse in any event effort to discover the information they need [3].

1.1 Web Mining

The term Web Data Mining is a method used to the crawl by numerous web resources to collect required knowledge, which enables an individual or a enterprise to promote their business, gone through the marketing dynamics, or the promotions on the internet, etc. There is a rising trend among companies, organizations and individuals to collect knowledge via web data mining to use that knowledge in their best interest.

A. Web Usage Mining

Web usage mining is the procedure of extricating helpful data from server logs i.e. users history. Web usage mining is the process of finding out what users are looking for on internet. Some users only focused on textual data, whereas some others might be interested in multimedia data as well. This technology is essentially determined upon the use of the web technologies which is helpful in increasing its usage. Web usage mining procedure includes the log time of pages. The world's biggest gateway like hurray, msn and so on., requirements a great deal of bits of knowledge from the conduct of their clients' web visits. Without this use reports, it will be hard to structure their adaptation endeavors. Usage mining has direct impact on businesses[5]. This is activity that contain automatic user access patterns discovery from various Web servers. As several organizations rely on the internet and using WWW to conduct businesses, the conventional approaches and methods for market analysis necessity to be returned to in this setting. Associations regularly produce and contain huge volumes of data in their everyday operations. Most of this knowledge is generally produced automatically through Web servers and contain in server access logs. Different wellsprings of client knowledge conclude referrer logs which contains knowledge about the referring pages for all page reference, and client enrollment or overview information gathered via tools for example CGI scripts [6].

- Web Server Data User logs are contain through web server and classically contain access time, IP address and page reference.
- Application Server Data Commercial application servers for example StoryServer, Weblogic, have important features to enable e-commerce applications to be constructed on top of them. A key element is the capacity to track various business sorts occasions and log them in application server logs.
- Application Level Data New events types can be characterized in an application, and logging can turned on for them — producing histories of these kinds of events. It must be noted, however, that numerous end applications need a combination of one or more of the methods applied in the above the groups [7].

B. Web Structure Mining

Web structure mining, one of three basic web mining groups for data, is a tool used to classify the relationship between Web pages connected through information or direct connection association. This structure information is discoverable through the procurement of the web structure composition by database methods for Web pages. This connection permits a search engine to pull information identifying with an inquiry question specifically to the connecting Web page from Web webpage the substance rests upon. This fulfillment happens by utilization of creepy crawlies filtering the Web destinations, recovering the landing page, then, connecting the information by reference connections to deliver the particular page including the desired knowledge. [7] Structure mining uses are minimized two main issues of the WWW because of its information quantity. The first of these issues is irrelevant search outcomes. Relevance of search knowledge become

misconstrued because of the issue that search engines often only permit for low precision criteria.

According to the web structural information type, web structure mining can be separated a into two different types:
1. Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a various location.
2. Mining the document structure: analysis of the tree-like structure of page structures to define HTML or XML tag usage [9].

C. Web Content Mining

Web content mining is the mining, integration and extraction of valuable information, data and information from Web page contents. Content mining is the text scanning and mining, graphs and pictures of a Web page to define the relevance of the search query content. This scanning is completed after web pages clustering by structure mining and gives the outcomes based upon the relevance level to the suggested query. With the massive quantity of knowledge that is presentable on the WWW, content mining gives the outcomes lists to search engines in highest relevance order to the keywords in the query. [8]

The web content mining is distinguished from the points of view: Information Retrieval present and Database View. R. Kosala et al. Summarized the research works complete for unstructured information and semi-structured information from knowledge retrieval view. It present that most of the researches use words bag, which depends on measurements about single words in confinement, to speak to unstructured content and assume single word found in the training corpus as types. For semi-structured information, each works use the structures of HTML inside the documents and few used hyperlink structure between the various documents for document representation. As for the database present, in order to contain the perfect information administration and questioning on the web, the mining dependably tries to gather the web site structure to transform a web site to become a database. This mining uses kind the principles and ideas of data mining and knowledge discovery to screen more particular data. The Web use as a, information gives is unfortunately more difficult than working with the static databases. There is a need for original solutions that are not depending on accessing the complete data on the outset [10]. Another significant aspect is the query presentation outcomes. Because of its huge size, a web inquiry can recover a large number of coming about website pages. Therefore important methodologies for exhibiting these huge results are important to assist a client with selecting the most intriguing substance.

1.2 Web Mining Application

Web mining broadens investigation much further by joining other corporate data with Web activity information. Viable utilizations of Web mining innovation are inexhaustible, and are in no way, shape or form the point of confinement to this innovation. Web mining apparatuses can be expanded and customized to answer almost any question. It can be applied in following areas:

- Web mining can give businesses, managerial insight into the visitor profiles, which help top management take planned actions accordingly[4].
- The company can find few particular measurements by Web Mining on the efficiency of their marketing campaign or marketing research, which will help the business to increase and align their marketing approaches timely.
- In the business world, structure mining can be entirely important in deciding the association between two or more business Web locales.
- This permits bookkeeping, client profile, stock, and demographic data to be corresponded with Web scanning
- The organization can distinguish the quality and shortcoming of its web promoting effort through Web Mining, and afterward make key alterations, acquire the input from Web Mining again to see the change.
- Search engine Google provides advanced and efficient searching capabilities[11].

2. RELATED WORK

Sequential pattern mining was first proposed by R. Agrawal and R. Srikanth [17]. He also proposed a comprehensive algorithm for sequential pattern mining GSP [18] to lessen down the search space for finding frequent sequences. Since there are several aalgorithms for sequential pattern mining. Closed sequential pattern mining is connected with mining of consecutive example and shut itemset mining. Consecutive example mining is utilized to find the total set of sequential patterns in a sequential database.

The author [19] has done the research on the information preprocessing in the web utilization mining. Information preprocessing is the imperative system in Web Usage Mining prior utilizing affiliation tenet mining as a part of the ensuing stages for instance ID exchange, investigation of way, successive example mining and affiliation guideline mining in the web log documents.

The creator [20] had complete the work on data preprocessing in web utilization mining. They displayed a novel calculation known as ─USIA. It finds the client and sessionDetails of ID. The same consumer is recognized with IP address and User ID help. If the request is from the same IP address, then algorithm concluded that the request is from the same user. The recognized session is based on time in and out period. When the requested page time is exceeded from 30 minutes, then the algorithm assumes to be started the new session again. In data cleaning phase unwanted data are removed to reduce the processing time of the algorithm. The process of data cleaning mainly used to improve the efficiency of the algorithm. For reference, the log file which has suffix .gif, .jpeg, .jpg is removed. So, this paper mostly focused on the user ID for the specific session and web pages series viewed through the user.

This author [21] intensive on group the consumer transactions with applying the clustering method. The transactions set in a group has some similarities, so we can simply recognized the client behaviour and the web site analyst can able to understand the consumer expectation and create the website consumer friendly. In another view point, create the website is more customized and more easy to use. The scientist utilized the example based grouping strategy to accumulation the comparative exchanges kind. Some measure is taken after to gather the exchanges, for instance {starting_time = morning, avg_time_page < 2, class = 3, total_time < 10 min} may be the behavioral pattenr for gathering the exchanges. The outcome may be the website pages of news, back, offer or email.

The creator [22] managed two unique gatherings sorts one is Web Clustering Groups which assembles the relative pages from the web server log records, the second is User Clustering Groups which amasses the client who alludes the same kind of website pages. Divisive Hierarchical Clustering Algorithm is utilized to aggregate the Web Log documents and User of comparative sort.Then the association rule mining with support and confidence measure is applied to each group to fine the relationship among them.

This author [23] focused on the first Web Usage Mining level known as Data Pre-processing and they suggested a new method for feature selection based on the Rough set Theory for Web Usage Mining. The issue in web Log Files is their unwanted and size information. This paper used two different algorithms Quick reduct and Variable Precision Rough Set Algorithm to classify the necessary information from the web log files, the actual feature selection procedure. The usage of k-means clustering algorithm is to segment the same patterns before using the above two different algorithms. So the algorithms are applied only to the similar items group to classify the feature selection. So, this method particular the optimal solution for removing the unwanted information in the web log files.

The author [24] mostly concerned about the data pre processing phase to eliminate the unnecessary information for example images, extra click events. Pattern discovery algorithms are used to remove the unwanted information from the web server log files. They taken the information from NASA website server log files and remove the unwanted data to progress the efficacy of the web log data analysing process. No particular data mining methods are applied to web log files after pre processing. That work is open for future research workers.

3. PROPOSED WORK

3.1 System Architecture

An architecture overview of the proposed method is given in Fig. 1. First, each users' web access actions of a website are recorded through the WWW server of the website and also stored into the Web Server Logs. All user access record holds the client IP address, request time, requested URL, user ID, HTTP status code, etc. After finding the Web Server Logs, the component of Sequential Pattern Mining is applied to mine the consumer access groupings from the Web Server Logs. At that point, the Pattern-tree Creation part builds the proposal model or Pattern-tree from the mined consecutive web access designs. Both the consecutive example mining and Pattern-tree development techniques are carried out off-line. The Pattern-

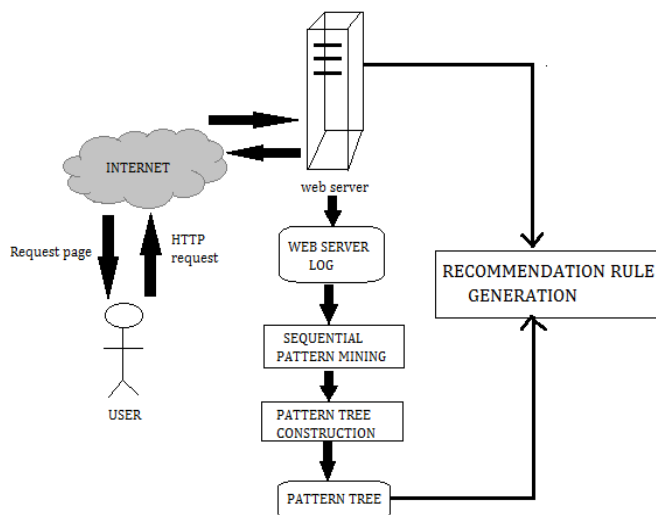tree can be updated commonly to incorporate novel access data.

Figure 1. System architecture

When a user visits the website, the user's HTTP requests in the present browsing session are recorded in order, and the present access sequence is constructed. All user accessing the website can be recognized applying his/her IP address. Matching the user's present access sequence from the proposal model of the Pattern-tree, the Recommendation Rules Generation segment will deliver suggestion rules. From the proposal standards, prescribed or "related" connections will be progressively embedded into the presently requested page.

3.2 Sequential Pattern Mining

In common, all web logs line contains the following key knowledge: date-timestamp, client IP address, user ID, requested URL, and HTTP status code. We describe web logs to be a sequences comprising web access events set from all user at the time of their corresponding session in timestamp ascending order. Preprocessing must be achieved on the web logs prior to using sequential pattern mining. The basic data preprocessing methods used here contain data cleaning, user identification, and session identification. Let's describe sequential pattern mining [25] as follows. Assumed a sequence database where all sequence is a transactions list ordered through transaction time with all transaction comprising a items set, find each sequential patterns with a user-specified minimum support, which is defined as the various data sequences including the pattern. Let E be a set of unique access events, which represents web resources accessed through users, i.e. web pages, URLs, topics or classes.

Sequential pattern is a itemsets set structured in sequence database which happens sequentially with a particular order. A sequence database is a ordered elements or events set, stored with or without a concrete notion of time. all itemset holds a items set which conclude the similae transaction-time value. While association rules specify intra-transaction relationships, sequential patterns represent the correlation between transactions. SPM [12] is the procedure that extracts various sequential patterns whose support exceeds a minimal support threshold. Moreover, sequential pattern mining helps to remove the sequences which reflect the most frequent

behaviors in the sequence database, which in turn can be construed as domain information for numerous purposes. To decrease the mosthuge number of sequences into the most interesting sequential patterns and to meet the various user requests, it is significant to use a minimum support which prunes the consecutive example with no hobby. It is clear that a consecutive example higher backing is favored for all the more intriguing successive examples. Consecutive example mining is utilized as a part of various spaces. SPM is utilized as a part of business associations to study client practices. Also, SPM is utilized as a part of computational science to investigate the amino corrosive change designs. SPM is additionally utilized as a part of the zone of web use mining to mine various web logs distributed on multiple servers.

Web personalization is any action that amends the data or services offered by a Web site to the requirements of a specific user or a set of users procuring benefit of the knowledge obtained from the user's navigational behavior and individual interests, together with the content and the structure of the Web site. The purpose of a Web personalization system is to supply users with the information they require, without expecting from them to ask for it clearly [19]. Web Recommender system is a kind of personalized web application which supplies considerable user value by personalizing a number of sites on the Web [20].
In our proposed work, sequential pattern algorithm has been explained like this:

Algorithm 1: Seq()

Step 1: Data is kept according to the user requirements.
Step 2: Pattern creation.
Step 3: Scan the database to find the pattern.
Step 4: Search for the patterns with 1-sequence and so on having spport more than threshold.
Step 5: Modification
  If (sub_seq occurs >3)
    Keep in the database.
  Else
    Reject.
Step 6: Return.

3.3 Pattern Analysis using Tree Creation

This analysis procedure would remove patterns or irrelevant rules that were created. They tend to remove the patterns or interesting rules from the output of the pattern discovery procedure. The most general pattern analysis form consists of an information query mechanism for example SQL or loads the usage information into a data cube in order to achieve OLAP operations. Visualization methods, for example graphing patterns or assigning colors to various values, can often highlight complete patterns or trends in the infomation. Numerous other mechanisms used for mining these patterns are

- *MWAP(Modified Web Access Pattern):* [21] This method entirely removes the requirement to engage the many reconstruction of intermediate WAP-trees the time of mining and considerably decreases execution time.
- *EXT-Prefixspan:* [22] This technique mines the patterns set but significantly decreases the efforts of candidate

subsequence generation. Prefix –projection procedure conclude in this technique substantially reduces the projected database size.

3.4 Pattern Tree Construction

A Pattern-tree model is proposed for storing sequential web access patterns efficiently, so that it can be used for matching with a customer's present access sequence and producing recommendation rules more effectively in the Recommendation Rules Generation component. In common, the sequential web access patterns set can be treated as a strings set over a finite alphabet E (the set of unique access events). All Pattern-tree node is labeled with a symbol (access event) from E and has a corresponding support value. Sometimes, the edges of a Tree are labeled nodes instead, but nevertheless refer to the same structure in either case. To create a Pattern-tree, we only essential to have one scan of entirely the sequential web access patterns.

Construction of Pattern tree
The construction of a pattern tree from the mined closed sequential web access patterns is described in this subsection. The constructed pattern tree is used in making the recommendation for a user's web access sequence. In general, the tree is used to store a set of strings but in regular tree, single character is stored in each node. By using tree, the tree can be even more compacted when compare with the regular tree.

Algorithm 2: Call_PTC()

Step 1: An empty root is taken for the construction of pattern tree.
Step 2: Set current node as the root node.
Step 3: Scan for the child node of the root node
Step 4: Maximum support is taken and current node is set.
Step 5: Otherwise, new node is created with support and set as current node.
Step 6: Return pattern tree.

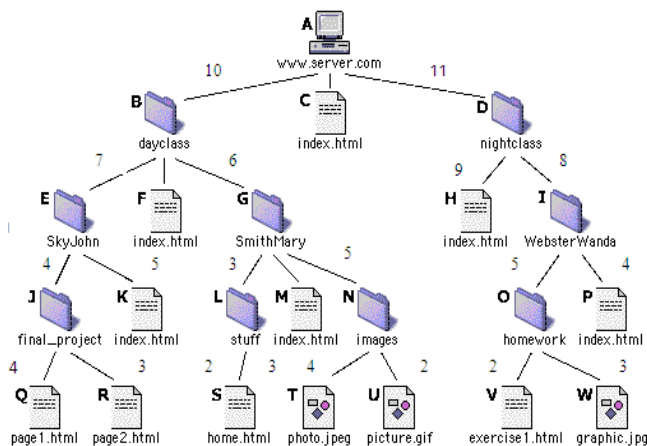**Figure 2:** Construction of Pattern Tree

3.5 Recommendation Rule Generation

The systems for web-page recommendation are based on collaborative-_ltering approaches. The underlying idea is to exploit the user-browsing logs (e.g. the click information), in order to identify users with similar interests and tastes. These recommender systems deal with very high levels of noise, since visiting a web page is not a clear indication of interest as renting a dvd or buying a book. Further, the web pages to recommend are not an input of the recommendation algorithm, and the systems have to discover the interesting pages to recommend.

The matching way will be not exist when the current access sequence length is longer than the Pattern-tree depth. Therefore, few initial items can be eliminated to create the current access sequence shorter than the Pattern-tree depth before the sequence matching process begins.

Algorithm 3: Call_Rec_rule_gen()

Step 1: Initialize root = null.
Step 2: Check if the search pattern given is greater than maximum length, then prune from the beginning else start searching.
Step 3: If search pattern is less than minimum length, then return empty root and no rule exist.
Step 4: for each item in search pattern
    If
        item has a child node, then set child node as current node
    Else
        Remove first item from search pattern and repeat from step 3.
Step 5: If current node has child nodes, then add the child node into the output patterns.
Step 6: Return output pattern.

The process of recommendation rule generation works n a particular algorithm that calls various algorithms within itself. The basic criteria behind the calling function are the whole process that is being taking place for the process.

3.6 Recommendation System Creation

The main process can be explained with the help of pseudo-code that represents the sequence of the processes taking place and the order in which the algorithms are being called.
The pseudo-code for the recommendation system proposed can be explained as below:-

Step 1: First clean the data and remove unnecessary data.
Step 2: Cleaned data is loaded into the database.
Step 3: Sequential pattern mining is applied resulting into frequent patterns.
    Call algo1();
Step 4: Tree is constructed for the frequent patterns.
    Call algo2();
Step 5: From the tree rules are generated.
    Call algo3();
Step 6: A new recommendation system is created.

4. PERFORMANCE EVALUATION

A. Evaluation Measures

Let S = a1a2…akak+1…an be a web access sequence. For the prefix sequence Sprefix = a1a2…ak (k ≥ MinLength), we generate a recommendation rule RR = {e1, e2, …, em} using the Pattern-tree, where all events are ordered by their support. If ak+1∈RR, we label the recommendation rule as correct, and otherwise incorrect. If there exists ai∈RR (k+1≤ i ≤ k+1+m, m>0), we deem the recommendation rule as m-step satisfactory. Otherwise, we label it as m-step unsatisfactory. Let R = {RR1, RR2, …, RRn} be a set of recommendation rules, where RRi (1≤ i ≤n) is a recommendation rule. |R| = n is the total number of recommendation rules in R. We define the following measures[23].

Definition 1: Let Rc be the subset of R that consists of all correct recommendation rules. The overall web recommendation precision is defined as

$$\text{Precision} = \frac{|Rc|}{R}$$

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

Definition 2: The net time utilized by the algorithm to process the code.

B. Experimental Results and analysis

The performance evaluation and the experimental results of the proposed recommendation system are presented in this section. The proposed recommendation system is implemented on MATLAB. The synthetic dataset is used for evaluating the performance of the web recommendation system. The used synthetic dataset has a collection of web access sequence and it is spitted into two parts:

1. *Training dataset:* It is used for designing the web recommendation system based on mined closed sequential patterns on it and
2. *Test dataset:* It is used to test the designed web recommendation system. At first, the pattern tree is constructed by using the training dataset and then, the proposed web recommendation system is evaluated with the test dataset by using the evaluation measures given below.

Result Obtained
Steps in the process of the implementation
Step 1 User is asked to input the search sequence
Step 2 User is asked to enter the suffix that he wants to search.
Step3 User is asked to enter m-step satisfactory recommendation rules to be searched.

These above steps are common for both the base and proposed papers. The results obtained are shown in tabular form below:

Table 1: Results of Time and Precision

For Base Code	Sequence	Suffix	Time	Precision
S1	2 5 5	5	55.3864	0.37500
S2	3 5 5	5	32.3012	0.37500
S3	9 9 9	9	19.8496	0.078947
S4	23 23 23	23	19.3520	0.714287
S5	22 23 23	23	19.6287	0.50000
For Propose Code	Sequence	Suffix	Time	Precision
S1	2 5 5	5	15.5634	0.428571
S2	3 5 5	5	22.7441	0.428571
S3	9 9 9	9	18.3102	0.108108
S4	23 23 23	23	19.0523	1.00000
S5	22 23 23	23	19.1210	0.750000

The graph shows in fig. 3 and fig. 4 that the proposed algorithm is high in precision as compared to the base algorithm. Also, the proposed algorithm is bettern in CPU time usage as compared to base algorithm.
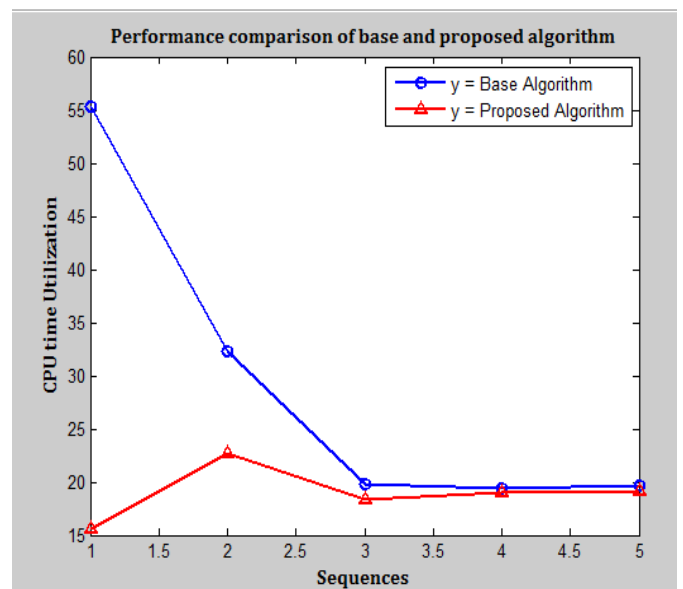

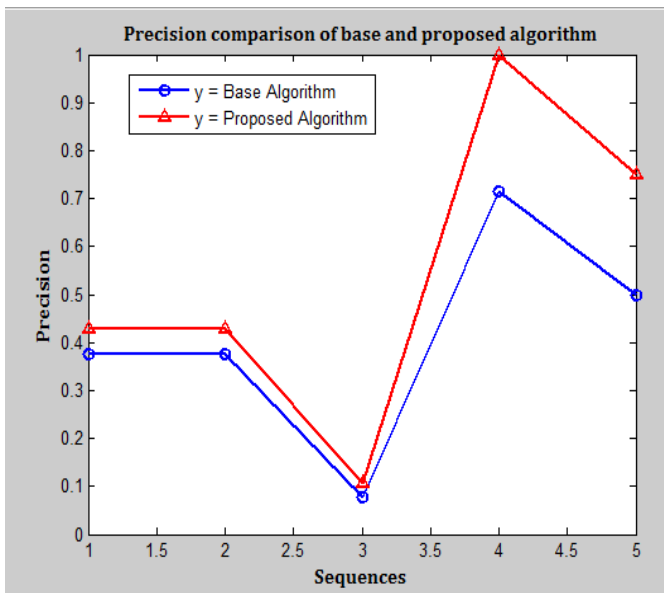**Figure 3:** Comparison of CPU time utilization

Figure 4: Comparison of Precision

5. CONCLUSION

The improved algorithm has been proposed which has various measures that are better as compared to the previously designed algorithm.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. In this research work, we have proposed a system for capturing the list of users to find the patterns which can be recommended to the users. Various factors, on which experiment has been performed, have shown betterment in every field. Hence, it is clear that the proposed approach works well and optimized which leads to the enhancement in the performance.

References

[1] Z. Maamar, G. Alkhatib and S.K. Mostefaoui, "Context-based Personalization of Web Services Composition and Provisioning," Proceedings of the 30th EUROMICRO Conference (EUROMICRO'04) IEEE 2004.

[2] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization,"ACM Transactions on Internet Technology, Vol 3,February 2003.

[3] S. Khonsha and M. H.Sadreddini, "New Hybrid Web Personalization Framework," IEEE 2011.

[4] Mary Garvin , "Data Mining and the Web: What They Can Do Together"

[5] B. Masand, M. Spiliopoulou, J. Srivastava, O. Zaiane, ed. Proceedings of "WebKDD2002 –Web Mining for Usage Patterns and User Profiles", Edmonton, CA, 2002.

[6] R. Kohavi, "Mining E-Commerce Data: The Good, the Bad, the Ugly", Invited Industrial presentation at the ACM SIGKDD Conference, San Francisco, CA, 2001.

[7] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, "Web Mining Concepts ,Applications and Research Directions".

[8] R. Kosala, H. Blockeel, "Web Mining Research: A Survey", in SIGKDD Explorations 2, ACM, July 2000.

[9] J. Srivastava, R. Cooley, M.Deshpande, P-N. Tan. " Web Usage Mining: Discovery and Applications of usage patterns from Web Data", SIGKDD Explorations, Vol1, Issue 2, 2000.

[10] T. Berners-Lee, R. Cailliau, A. Loutonen, H. Nielsen, and A. Secret. The World- Wide Web. Communications of the ACM, 37(8):76- 82, 1994.

[11] R. Cooley, B. Mobasher, J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, 1(1), 1999.

[12] Rakesh Agrawal, SaktiGhosh, Tomasz Imielinski, BalaIyer, and Arun Swami, An Interval Classier for Database Mining Applications", VLDB-92 , Vancouver, British Columbia, 1992, 560-573.

[13] Pradnya A. Shirsath, Vijay Kumar Verma, "A Recent Survey on Incremental Temporal Association Rule Mining", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue-1, June 2013.

[14] Yiyu Yao, "Rough Set Approximations: A Concept Analysis Point Of View", University of Regina, Regina,Saskatchewan, Canada, 2015.

[15] En-Bing Lin and Yu-Ru Syau, "Comparisons between Rough Set Based and Computational Applications in Data Mining", International Journal of Machine Learning and Computing, Vol. 4, No. 4, August 2014.

[16] Dewang Rupesh, Agarwal Jitendra. "A New Method for Generating All Positive and Negative ssociation Rules":International Journal on Computer Science and Engineering,(2011) ,Vol. 3, pp:1649-1657.

[17] Abdoljabbar asadi, Mehdi Afzali. "Providing a newmethod for detecting positive and negative optimal performance association rules in very large databases using Binary Particle Swarm Optimization": The sixth Iran Data Mining Conference / IDMC, Dec 01,01 / 2102, Tehran, Iran,(2102).

[18] Abdoljabbar Asadi, Mehdi Afzali, Azad Shojaei, Sadegh Sulaimani. "New Binary PSO based Method for finding best thresholds in association rule mining". Life Science Journal; 9(4).pp:260 - 264,(2012).

[19] Masseglia F., Poncelet P., and Teisseire M. (1999). Using data mining techniques on web access logs to dynamically improve hypertext structure. In ACM SigWeb Letters, 8(3): pp. 13-19.

[20] Zhang Huiying, Liang Wei.An (2004). Intelligent Algorithm of Data Pre-processing in Web Usage Mining. In Proceeding of the 5th World Congress on Intelligent Control and Automation. pp. 15-19. Hangzhou, P.R. China.

[21] Yinghui Yang and Balaji Padmanabhan. (2005). GHIC: A Hierarchical Pattern-Based Clustering Algorithm for Grouping Web Transactions. IEEE Transactions on Knowledge and Data Engineering, Vol 17, No. 9.

[22] Yi Dong, Huiying Zhang and Linnan Jiao. (2006). Research on Application of User Navigation Pattern Mining Recommentation. In Proceeding. of the 6th World Cogress on Intelligent Control and Automation. Dalian, China.

[23] Hannah Inbarani H., Thangavel K., and Pethalakshmi A. (2007). Rough Set based Feature Selection for Web Usage Mining. International Conference on Computational Intelligence and Multimedia Applications.

[24] Suneetha K. R., and Krishnamoorthi R. (2009). Identifying User Behavior by Analysizing Web Server Access Log File. IJCSNS International Journal of Computer Science and Network Security, Vol 9, No.4.

[25] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation or mining World Wide Web browsing patterns", Journal of knowledge and Information Systems, Vol. 1, No.1. 1999.

[26] David E Goldberg, Holland John H, "Genetic Algorithms in Search, Optimization and Machine Learning", Reading, MA:Addison-Wesley.

[27] Dewang Rupesh, Agarwal Jitendra. "A New Method for Generating All Positive and Negative Association Rules":International Journal on Computer Science and Engineering,(2011) ,Vol. 3, pp:1649-1657.

[28] Jatin D Parmar, and Sanjay Garg,(June, 2007)"Modified web access pattern (mWAP) approach for sequential pattern mining," INFOCOMP Journal of Computer Science., pp. 46-54.

[29] S.Vijayalakshmi V.Mohan, S.Suresh Raja, (2009) "Mining Constraint-based Multidimensional Frequent Sequential pattern in Web Logs," European Journal of Scientific Research., Vol.36, pp 480-490.

[30] Utpala Niranjan1, Dr.R.B.V. Subramanyam2, Dr.V.Khanaa, "An Efficient System Based On Closed Sequential Patterns For Web Recommendations", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 4, May 2010.

Authors

Dr. Sanjiv Sharma currently works as an Asst. professor in Department of Computer science Engineering and Information Technology at Madhav Institute of Technology and Science, Gwalior, India. He has 9 years of teaching and research experience. He has received his B.E. from Pt. Ravishankar University, Raipur, and M. Tech. from RGPV, Bhopal, and Ph. D from Bansthali Vidhya Peeth, Jaipur, India. His area of research is Network security, Data Mining and Ad-hoc and wireless sensor Networks.

Kuldeep Singh Rathore currently pursuing M.Tech in Computer Science and Engineering from Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India under the guidance of Dr. Sanjiv Sharma. He has received his B.E in Computer Science and Engineering from RGPV, Bhopal Madhya Pradesh, India. His area of interest is Data Mining, Web Mining and its sub field that is Web Usage Mining.