

A Review of Web Personalization and Recommendation for Making Decision using Sequential Pattern

Kuldeep Singh Rathore¹ Sanjiv Sharma²

¹Research Scholar ²Assistant Professor

^{1,2}Department of Computer Science & Engineering

^{1,2}MITS Gwalior, India

Abstract— Data mining is the field that has been used to mine related data in an appropriate manner. The development of web has generated a big issue for directing the client to the website pages in their area of interest. Accordingly, just option is to capture the intuition of the client and provide them a list of recommendation. Web usage mining is used to mine knowledge based on the user log. Sequential mining is the using data mining methods process to a sequential database for the discovering purposes the correlation relationships that occur among an ordered events list. The recommendation is web usage mining application. Web personalization is any action that amends the services or data offered through a Web site. Web personalization goals is to recommend the most related resources to a user at the time of a particular time period through analyzing the periodic access patterns of the user from the web usage logs. This paper gives a brief description on data mining, and the applications of web mining, concepts algorithm and web personalization and previously work done.

Key words: Data Mining, Web Usage Mining, Sequential Pattern Mining, Web Personalization, Recommender System

I. INTRODUCTION

Data mining [1] is the discovering insightful process, interesting, and novel patterns, as well as predictive, understandable, and descriptive models from big-scale data. The objective of data mining is to recognize legal new, potentially helpful, and reasonably correlations and patterns in presenting data.

Data Mining trends can be ordered into two classifications, Descriptive Mining and Predictive Mining [2]. The Descriptive Mining strategies, for patterns and Clustering, are utilized to discover human-interpretable patterns that depict the information. In data mining Sequential Pattern Discovery, Association Rule Discovery, web mining concept is used to remove knowledge from unstructured data. The main purpose of web mining is to extract data from the web explicitly. Web mining has various sub tasks they are Resource finding, Information Selection and pre-processing, Generalization, Analysis Resource Finding is the task of retrieving intended Web documents. Knowledge Selection and pre-processing is automatically selecting and pre-processing specific information from retrieved from resources of web. Generalization is used for selecting and pre-processing specific information explicitly from retrieved web resources. The validation and interpretation of mined patterns analysis performed. In next section discuss web mining their types and concept of web usage mining, approaches, algorithms, and previously done related work.

II. WEB MINING

Web Mining is used to remove knowledge from the raw unstructured information. The emerging web mining techniques aims area at discovery and removing irrelevant information that is not directly seen in web related information, in specific in text documents published on the Web. Web mining is vastly divided into three categories:-

- Web usage mining
- Web content mining
- Web structure mining.

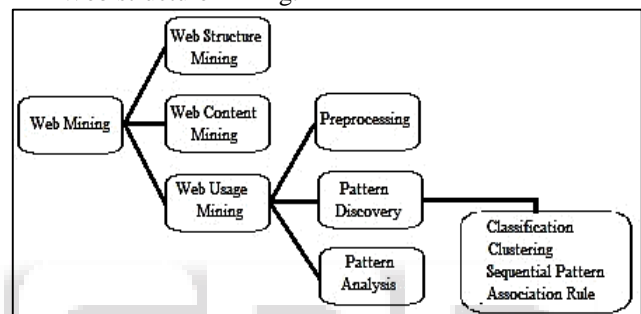


Fig. 1: Classification of web mining

Web usage mining gives the web site design support, supplying personalization server and different business creating decision, etc. Web content mining is the procedure of mining information from the documents content or their descriptions. Web document text mining, resource discovery based on idea indexing or agent; based technology may also fall in this category. Web structure mining is the inferring knowledge process from the WWW organization and links between references and referents in the Web.

Web usage mining is used to mine knowledge based on the user log. WUM is the method of applying information mining strategy to the usage patterns discovery from Web data, targeted towards numerous applications.

The usage data collected at the various sources will represent the navigation patterns of various segments of the complete the traffic from web, ranging from single user, and single site browsing behavior to multi-user, multi-site access patterns.

A. Concept of Web Usage Mining

Discovery of significant patterns from data generated through client-server transactions on one or more web servers. Classical Sources of Data:

- Automatically generated data stored in referrer logs, server access logs, agent logs, and client-side cookies.
- Electronic commerce and product-oriented user events (e.g. shopping cart changes, ad or product click-through, etc).
- User ratings and/or User profiles.
- Meta-data, page attributes page content, site structure.

B. Approaches of Web Usage Mining

- Data Collection: accumulation of data is the essential level of web usage mining, the data authenticity and integrality will straightforwardly influence the accompanying works effectively carrying on and the last trademark administration's quality proposal. Along these lines it must utilize exploratory, progressed and sensible innovation to assemble various data. At current, towards web usage mining technology, the essential data origin has 3 sorts: middle data, server data and client data.
- Data Preprocessing: Couple of databases is conflicting, deficient and including clamor. The data pretreatment is to bear on a unification change to those databases. The result is that the database will to end up coordinate and steady.
- Knowledge Discovery: Use statistical technique to carry on the mine and analysis the pretreated information. We might decide the individual or the user neighborhood's interests then construct interest model. At current the typically utilized machine learning approaches for the most part have clustering, classifying, the connection revelation and the request model disclosure. All system has its own shortcomings and excellence, yet the very productive strategy for the most part is clustering and classifying at the present.
- Pattern Analysis: Pattern Analysis Challenges are to filter uninteresting knowledge and to interpret and picture the most interesting examples to the customer. In the first, erase the less significance models or rules from the most intrigued model storage facility; Another use OLAP innovation thus on to bear on the complete analysis and mining ; over once more, let found data or information be unmistakable; Finally, give the trademark office to the E- commerce website.

III. SEQUENTIAL PATTERN MINING

Sequential pattern mining is to mine the frequently occurred patterns related to the items order in a sequence database. The discovered frequent data is useful for broad application, such as retail business, disease treatments, market analysis, etc. The task of sequential pattern mining is helpful for different applications, including market analysis, decision support, and business administration. One vital issue is to find continuous sequential patterns. Also, the vast majority of the past works have concentrated on the order of the times. Sequential pattern mining is significant task in data mining. Algorithms of sequential pattern mining established so far provide higher presentation for short sequences but these algorithms are ineffective at mining long sequences. Sequential sample mining is large information mining subject, a further association rule mining promotion, and it is also extensively applied [3]. Algorithms of consecutive pattern mining address the finding issue the current incessant arrangements in a give database [4]. Consecutive pattern mining is most firmly identified with mining of association rule, aside from that developments are connected through time [5]. Successive patterns show the connection between's exchanges while organization standard presents intra exchange connections. In organization rule mining, the mining results are about which questions are united

customarily, those things need to originate from the identical transaction. In the meantime the sequential pattern mining results are about which things are purchased in a different request through the same customer, with those things originating from various transactions [6]. Sequential patterns can offer supervisors some assistance with defining which things are acquired one after a further in an arrangement [7], or to analyze browsing, landing pages orders in a Web site [8] and then some. Sequential pattern mining is normally characterized as disclosure the complete incessant set subsequences in an arrangements set [9]. Sequential pattern is a thing arrangement sets that as often as frequently happen in a specific request; each things in the same thing set should contain the same exchange time worth or inside of a period gap. All arrangement compares to a transiently ordered events list, where all occasion is items collection (thing set) happening all the while.

The sequence Data Mining [10] thought was initially presented. The essential issue was initially presented in the market analysis context. It objective to examples of recover incessant in the items groupings obtained through clients by time requested exchanges. Later in its application was reached out to complex applications for instance network detection, telecommunication, DNA research, etc. Various algorithms are there. The first one was Apriori algorithm, which was put forward through the founders themselves. After extra scalable algorithms for complex applications was developed .E.g. PrefixSpan, GSP, Spade etc. The area underwent developments on the grounds that its introduction in a short span.

A. Basic Concepts of Sequential Pattern Mining

- 1) Let $I = \{x_1, \dots, x_n\}$ be a things set, all maybe being associated with a attributes set, for occurrence esteem, calling detachment, period, advantage, etc. The value on property an of thing x is meant through $x.A$. An item set is a non-unfilled things subset, and an item set with k things is known as k -item set.
- 2) An arrangement $\alpha = \langle X_1 \dots X_l \rangle$ is an asked for item sets list. An item set X_i ($1 \leq i \leq l$) in a course of action is known as trade, a term started from analyzing consumers' " shopping groupings in a trade database. A trade X_i may have a particular trademark, time-stamp, meant through $X_i.time$, which the time that was registered where all transactions was executed. For a gathering $\alpha = \langle X_1 \dots X_l \rangle$, we acknowledge $X_i.time < X_j.time$ for $1 \leq i < j \leq l$.
- 3) The diverse transactions process that is used in an arrangement is known as the succession length. A progression with length l is known as an l -gathering. For a l -succession α , we have $len(\alpha) = l$. Furthermore, the i -th itemset is implied by $\alpha[i]$. An item can occur at most one time in an item set, however can happen diverse times in different item sets in a progression.
- 4) An arrangement $\alpha = \langle X_1 \dots X_n \rangle$ is known as a subsequence of another sequence $\beta = \langle Y_1 \dots Y_m \rangle$ ($n \leq m$), and β a super-succession of α , if there exist whole numbers $1 \leq i_1 < \dots < i_n \leq m$ such that $X_1 Y_{i_1}, \dots, X_n Y_{i_n}$.
- 5) An arrangement database SDB is a 2-tuples set (sid, α) , where sid is a gathering id and α a progression. A tuple (sid, α) in arrangement database SDB is said to

contain a gathering γ if γ is a subsequence of α . The distinctive tuples in a progression database SDB including gathering γ is known as the support of γ , demonstrated with $\text{sup}(\gamma)$. Given a positive number min_sup as far as possible, a course of action γ is a progressive case in gathering database SDB if $\text{sup}(\gamma) \geq \text{min_sup}$. The Sequential pattern mining issue is to find the complete sequential cases set with respect to a given gathering database SDB and a threshold min_sup .

In light of these conditions sequential pattern mining can be isolated extensively into two sections:

- Apriori based (GSP, SPADE, SPAM)
- Pattern development based (Free Span, Prefix Span)

B. Apriori Based Algorithm

The Apriori and the Aprioriset [10] are the basis for an algorithm that depend on apriori property and use the join concept of Apriori-generate algorithm to generate candidate sequences. The property of apriori states that "All nonempty frequent item set subsets must also be Frequent". It is also define as ant monotonic (or downward-closed), in that if a series cannot cross the minimal help scan, its complete super sequences will also fail the test. Apriori-based algorithm Key features are: [11]

- Breadth-First Search: Apriori-based algorithms are define as breath-first search algorithms because they concept each the k-sequences, in algorithm of kth iteration, as they traverse the search space.
- Generate-and-Test: This service is used through the earliest algorithms in sequential sample mining. Algorithms that depend upon this option only present an inefficient pruning technique and produce an explosive various candidate sequences and then test all one by one for satisfying few users identified constraints, consuming a lot of memory space in the previous mining stages.
- Multiple Scans of The Database: This scanning of usual database to test whether or not an extended generated candidate sequences list whether frequent or not. This process is undesirable from most apriori-based algorithms and includes the condition of processing time and I/O cost.

C. Pattern-Growth Algorithms

Soon after the mid-1990s Apriori-headquartered approaches, pattern progress-method emerged within the early 2000s, as an answer to the generate-and-test issue. The basic concept is to avoid candidate new release step altogether, and to focus the quest on a constrained initial database portion. The search space partitioning feature performs a significant role in pattern-growth. Pattern growth-based algorithm having various key features:

- Search Space Partitioning: Search space partitioning performs partitioning of the big candidate sequences produces search space for efficient memory management. There are various ways to partition the search space. Once the search space is divided, smaller divisions can be mined in parallel. Advanced methods for search space partitioning include projection databases and conditional search, called as split-and-project methods.

- Tree Projection: Tree projection commonly accompanies algorithms of pattern-growth. Here, algorithms implement a data structure of physical tree representation of the search space, which is then traversed depth-first or breadth-first in frequent sequences search, and pruning is based on the apriori property.
- Depth - First Traversal: The DFS of space makes a major variety in execution, furthermore helps in the early applicant successions pruning and shut arrangements mining [11]. The essential explanation behind this execution is the way that profundity first traversal makes utilization of far less memory, additional coordinated inquiry space, and in this manner grouping era of the less hopeful than broadness first or post-request which are utilized through couple of ahead of schedule algorithms.
- Candidate Sequence Pruning: Pattern-growth algorithms attempt to utilize a data structure that allows them to prune hopeful groupings ahead of schedule in the mining method. This result in right presentation of small inquiry space and kept an additional coordinated and narrower search process timely.

IV. APPLICATION AREAS

Sequential pattern mining is utilized as a part of a field's assortment. The essential Sequential pattern mining point is to find the continuous subsequences in a dataset. Sequential pattern mining has numerous applications in the field of solution, DNA sequencing, in the medical field, sequential patterns for signs and illnesses present by patients recognize strong reaction or sickness relationship that can be a huge wellspring of data for restorative analysis of that data and various preventive and useful medicine. The sequential pattern mining of the medicinal subtle element can reveal the amount of time a disease can take to achieve an ailment after it has polluted somebody. Time-interim progressive examples engage authorities to keep their patients from getting the opportunity to be polluted through different diseases. They can in like manner help experts to utilize sound judgment when diagnosing their patients' diseases [12].

DNA sequencing, the DNA grouping design contains of four letters in order (A, G, T, and C) sorts. The example additionally finishes up a crevice in the DNA grouping. Through discovering the DNA arrangement with hole, the obscure successions can be grouped and found into its relating DNA family and this can be utilized for further research as a part of biological analysis [13].

Sequential Pattern mining is valuable in the Bioinformatics field for foreseeing rules for relationship of specific parts in qualities and for protein function prediction. It is additionally utilized as a quality expression part, examination and for protein fold recognition [14].

Sequential Pattern mining can be utilized as a data exchanges field part for mining of gathering patterns improvement of mobile client, data and for envisioning future versatile client region for range based organizations. It can in like manner be used for customer behavior figure and for digging patternss supportive for M-commerce [14].

Sequential pattern mining can be conceivable on a website page. Web log analysis [15], the exploring client conduct can be expelled from part records or log files. For this condition the arrangement could be Web Pages succession went by through clients on a webpage. By then algorithm of sequential pattern mining could utilize sequential pattern mining to find site pages plans that are ordinarily gone by through customers. The site could use these examples to make proposals to the customer, for examples, recommended joins. Sequential data are gathered intermittently from web server logs, online transaction logs, execution estimation. This presentation helps in hunting down a particular quality or occasion and it might likewise help with the analysis of the frequency of different occasions or sets of related occasions. Business endeavors make sequential pattern mining use to study the client practices. It is likewise utilized as a part of the analysis of telecom network analysis and system execution. Sequential pattern mining is utilized to dissect the change examples of different amino acids in computational biology.

V. WEB PERSONALIZATION AND RECOMMENDER SYSTEM

Web personalization is any activity that corrects the services or data offered through a Web site to the need of a specific client or a clients securing set point of interest of the data acquired from the shopper's navigational conduct and individual hobbies, together with the substance and the Web webpage structure [16]. The Web personalization framework intention is to supply clients with the learning they require, without anticipating from them to request it unmistakably [17]. The web Recommender system is a customized web application kind which supplies extensive client value through customizing different locales on the Web [18].

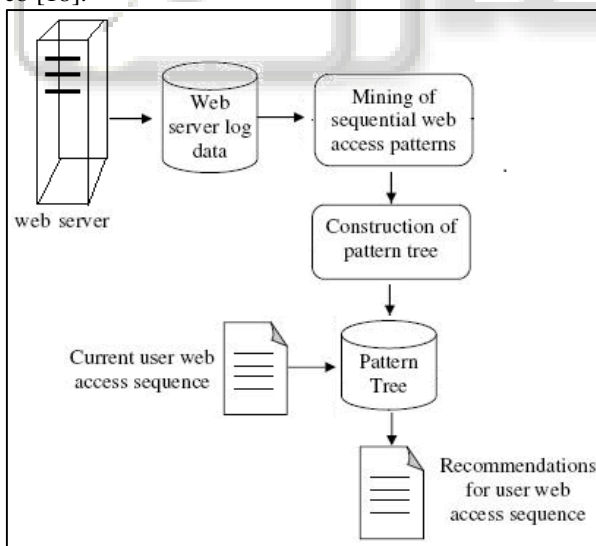


Fig. 2: Web Personalization Framework

Web personalization is a methodology, marketing software, and a work of art. Personalization require unequivocally or verifiably gathering guest learning and utilizing that data in your substance conveyance system to the work what information you present to your clients and how you show it. Legitimately accomplished, personalization of the voyager's mastery makes his time on your site, or in your application, extra engaging with and productive.

Personalization can likewise be esteemed to you and your organization, since it drives preferred industry results for instance developing guest reaction or empowering client maintenance. Unfortunately, personalization for its own particular purpose can possibly development the your site interface many-sided quality and drive wastefulness into your design. It may even trade off the you're promoting data viability or, more terrible, hinder the client's experience. A few organizations are willing to drawback their center information for the purpose of a some trap website pages. In opposition to famous conviction, personalization doesn't contain to take the altered substance entrances structure, promoted in the mid-to-late 90s through snap.com furthermore My Yahoo! Nor does personalization require live-in consultants or expensive applications. Personalization can be as explicit or as downplayed as you need it to be. It's a tired historical yarn, however in the event that you happen to would like to put in power a web personalization approach, the first and most huge level is to develop and add to your business points and need. It is critical to detail what it is you would like to do and, from that data, build up a how you understand from execution thought. You may be shocked to find that it won't require the majority of another year's financial plan to accomplish advantageous results. Web personalization can be available as an interdisciplinary field that closes different exploration areas from buyer displaying [14], social networks[19], web data mining [8,13,19], human-machine collaborations to Web usage mining [13]; Web usage mining is a case of strategy to remove log files containing learning on shopper route keeping in mind the end goal to sort clients. Other techniques of knowledge retrieval depend on archives categories' selection [13].

The Web personalization technique can be isolated into four different stages [13, 20]:

- Web Data Collection: Implicit data finish up past click/activities streams as recorded in Web server logs and/or through session or treats following modules. *Explicit* data by and large originates from rating polls and *registration* forms. Extra data for instance demographic and application information, (for example, electronic trade exchanges) can likewise be utilized. In few cases, Web substance, structure, and application information can be included as additional information sources, to reveal more insight into the stages.
- Web Data Preprocessing: Information is usually pre-prepared to place it into an arrangement that is suitable with examination technique to be utilized as a part of the another step. Preprocessing might finish up cleaning irregularities information, sifting through unessential learning as indicated by the analysis goal (sample: consequently created need to installed representation will be recorded in web server logs, even by they include little information about buyer premiums), and finishing the missing connections in inadequate click by paths.. Most very, novel sessions required to be distinguished from the different solicitations, in view of a heuristic, for instance demands beginning from location of indistinguishable IP inside of a given time period

- Analysis of Web Data –Likewise called as Web Usage Mining [18,22], this step applies Data Mining and machine learning approaches to manage find fascinating use plans and factual connections between's customer social events and pages. This stride every now and again results in programmed client profiling, and is traditionally used logged off, with the goal that it doesn't include a weight the net server.
- Decision making/Final Recommendation Phase – The last stage in personalization makes results utilization of the prior examination venture to offer proposals to the individual. The suggestion methodology traditionally finish up producing dynamic Web content on the fly, for instance adding hyperlinks to the last net website page asked for through the client. This can be expert applying a Web innovation assortment alternative for instance CGI programming.

Web recommender systems have drawing attention regarding extent as and a proper strategy to counter learning over-burden and help the shoppers of the Web data space to find what they require faster [19]. Web recommendation systems [20] are significant in educating the clients to the objective pages. On the other distinctive hand, utilizing proposal frameworks to introduced Web destinations required extensive work and data [21]. There exist various ways to deal with fabricate such frameworks. Among these methodologies, the web suggestion framework working by web access log is an acknowledged procedure [22]. The info to the framework will be the web server log data.

VI. DECISION MAKING IN WEB PERSONALIZATION

For performing the task of decision making the concept of Decision Support System is used. (DSS) is a tool used to improve the process of decision making in complex systems. A DSS can covers from a system that replies easy queries and allows a subsequent right decision to be made, to a system that applied artificial intelligence and provides detailed query across all the spectrum of related datasets in database. From the most important applicable areas of DSS are those complex systems that directly answer questions, in particular high level what-if scenario modeling. During previous decade there was a transition to decision support using data warehouses. In data warehouse scenario it is more controlled and hence more durable and reliable for decision support than the previous methods.

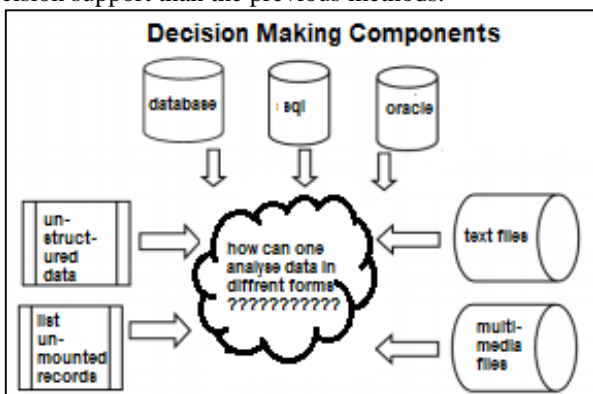


Fig. 3: Components of Decision Making

VII. RELATED WORK

Sequential pattern mining is computationally testing in light of the fact that such mining might create and/or test a combinatorial explosive different moderate grouping. Various novel calculations are proposed for instance Apriori ALL, Apriori, GSP, SPAM, SPADE and Prefix Span.

Jian Pei, et al. [23] have completed a legitimate prepared on the consecutive examples mining in the Gazelle information set from Blue Martini and an example development system had been proposed for the viable and adaptable successive examples mining. Rather than grouping designs refinement like in the apriori-like furthermore rather of competitor emphasis and-analysis technique for instance in GSP, a divide and-conquer strategy known as the example development technique, Prefix Span is elevated which turns out to be a powerful example development calculation for mining general examples without hopeful cycle. Prefix Span recursively undertakings a chain database into an accumulation of littler anticipated succession databases and development consecutive examples in all anticipated database through investigating just close-by continuous sections. The complete consecutive example set is mined and significantly diminishes the hopeful subsequence era endeavors.

Thomas. Rincy. N and Yogadhar Pandey [24] watched the execution assessment, pattern dataset present that the SPAM method performs quite enhanced and has an enhanced versatility than Prefix Span in execution time terms while in memory use terms, the system obviously shows that SPAM has stable memory utilization than PrefixSpan for every base bolster values. Prefix Span algorithm is executed with pseudo-projection strategy and still by watching the execution assessment pattern it unmistakably introduce that SPAM can be speedier on thick and meager Datasets, likewise the memory utilization is enduring as in correlation with PrefixSpan which is conflicting to the established angle. SPAM it typically expends additional memory than PrefixSpan and SPAM is speedier on thick datasets with long examples and less successful on other dataset furthermore it devours additional memory.

Mahdi Esmaeili and Fazekas Gabor [25] theoretically have present three consecutive examples sorts and few of their properties. These models fall into three classes are known as factually example, periodic pattern and approximate pattern. Periodicity can be finished periodicity or fractional periodicity. In complete periodicity procedure, unsurpassed point adds to the cyclic time arrangement conduct. In qualification a while components in incomplete periodicity make a commitment to the cyclic behavior of a period succession. This mannequin of test is so rigid. The knowledge gain can be utilized as a novel metric that help us to find the surprising patterns.

Niti Desai and Amit Ganatra[26] had performed a hypothetical and diversion study on different algorithms of sequential pattern mining. They suggested that PrefixSpan is a convincing case creating strategy since it beats GSP, Free Span and SPADE. They present that the Prefix Span Algorithm is extra intense with see to running time, house use and versatility than Apriori masterminded calculations. By far most of the current SPM calculations wear down target measures Confidence and Support. Their trials

demonstrate that the rate guideline time diminishing is high if there ought to be an event of interestingness measures lift. They moreover elucidated that interestingness measures use can incite make the case extra fascinating and can provoke indentify emerging patterns.

Ching-Yao Wang [27] has proposed algorithm for sequential pattern mining in light of the incremental mining idea. This algorithm utilizes the Pre-Large grouping idea to minimize the required for rescanning the first databases. Through utilizing the upper bolster limit and lower bolster edge it characterizes the Pre-Large arrangement that execute as hole to oppose the grouping development from expansive too little and from little to extensive. This algorithm does not perform the rescanning of the database until the novel buyer arrangement is added. When database size get bigger, the quantity of novel exchanges perform before the database rescanning require likewise develop.

Chi-Yao Tseng [28] have proposed a general model for consecutive example with the evolving database, while the data in the database can be settled, included or uprooted. Likewise, they displayed the progressive algorithm known As PISA which is stands for innovative mining of Sequential example which identifies the sequential pattern in altered time enthusiasm for dynamic way. The time of hobby is the time period, persistently advancing with time passes by. In PISA algorithm, to productively keep up the present data sequences it uses a dynamic grouping tree. It discovers outs the entire arrangement of up and coming successive example and wipe out old data and example according to require. The successive example size tree delivered was relying upon the length of the time window period. So that effectively minimizing the memory need by calculation that is not exactly the memory need by different strategies.

Jiaxin Liu [29] have proposed an data storage structure, called frequency sequence tree, and gives the time strategy for the incessant arrangement tree called con FST. At the root node of this continuous arrangement tree set away the regular grouping of nodes to support tree and the path from the hub to the any outer node addresses a progressive case in the database. The consecutive example whose reinforce meets the nonstop progression tree bolster farthest point is secured in a typical plan tree, so as changed the limit structure find the complete successive example without mining the complete unique database.

Vincent Shin-Mu Tseng [30] have proposed the standard development, the procedure for mining the successive tenets same for different sequences. Dissimilar to the distinctive algorithms principle development depends on the example development technique for finding consecutive example decides such that it can be versatile and better. They performed test of the standard developing with other few calculation on general society datasets. It found that the guideline development obviously beats the other diverse calculations, for these datasets under settled limit and low support.

Jiaxin Liu [31] have recommended that the sequence tree structure in light of the anticipated database, suggested as sequence tree, for the sequence tree development they proposes the steeps algorithm. Succession Tree was the data storage structure. It is comparable in structure to the prefix tree. Be that as it may, it stores each

the grouping in the first database. The way from the root node to any leaf node is a grouping in the database. The fundamental characteristics of the grouping tree make it legitimate for the increase pattern mining. From the experimental analysis demonstrate that the increment mining technique of back to back sample which uses the progression tree as the capacity structure for arrangement design performed better as comparison to prefix span in memory utilized cost on condition that reinforce edge was littler. To take the dynamic way of data addition and deletion.

VIII. CONCLUSION

This paper present a survey on web personalization in view of sequential mining algorithms and related work. Web personalization is a methodology, and a marketing software. Decision making with web personalization using support system also taken into account. A study on the sequential patterns mining in huge databases furthermore built up an example development procedure for versatile and effective mining of successive examples. As opposed to refinement of the from the prior, for example, contender time and-test methodology, for instance, GSP [23], we advance a partition and-overcome system, known as case improvement strategy, which is FP-growth extension [9], a compelling pattern-growth algorithm for mining successive examples without applicant era. There are different fascinating issues that required to be focused further, for example mining shut and maximal sequential patterns, and so on.

REFERENCES

- [1] Han J and Kamber Data Mining: Concepts and Techniques. Second edition Morgan Kaufmann Publishers.
- [2] Pang Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data mining", (2009).
- [3] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang- Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", Volume 1, Issue 2 - page 12, ACM, 2000.
- [4] Chintan R. Varnagar, Nirali N. Madhak, Trupti M. Kodinariya, Jayesh N. Rathod, "Web Usage Mining: A Review on Process, Methods and Techniques", IEEE, 2013.
- [5] Olfa Nasraoui, Maha Soliman, Esin Saka, "A Web Usage Mining Frame work for Mining Evolving User Profiles in Dynamic WebSites", IEEE transactions on knowledge and data engineering, vol. 20, no. 2, February 2008.
- [6] V.Chitraa, Dr.Antony Selvadoss Thanamani, "A Novel Technique for Sessions Identification in Web Usage Mining Preprocessing", International Journal of Computer Applications, Volume 34- No.9, November 2011.
- [7] Ankita Kusmakar, Sadhna Mishra "Web Usage Mining: A Survey on Pattern Extraction from Web Logs", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September, 2013.
- [8] Sanjeev Dhawan, Swati Goel, "Web Usage Mining: Finding Usage Patterns from Web Logs", American

- International Journal of Research in Science, Technology, Engineering & Mathematics, 2013.
- [9] G T Raju, Nandini N, "Preprocessing of Web Usage Data for Application in Prefetching to Reduce Web Latency", International Journal of Electrical & Computer Sciences IJECS- IJENS Vol: 14 No: 04, August, 2014.
- [10] Yogish H K, Dr. G T Raju, Manjunath T N, "The Descriptive Study of Knowledge Discovery from Web Usage Mining", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
- [11] Chetna Chand, Amit Thakkar, Amit Ganatra, "Sequential Pattern Mining: Survey and Current Research Challenges", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.
- [12] Yen-Liang Chen, Mei-Ching Chiang, Ming-Tat Ko Discovering time-interval sequential patterns in sequence databases, Expert Systems with Applications 25 (2003) 343– 354.
- [13] Syeda Farzana, Byeong-Soo Jeong-A Fast Contiguous Sequential Pattern Mining Technique in DNA Data Sequences Using Position Data.
- [14] Manish Gupta, Jiawei Han- Applications of Pattern Discovery Using Sequential Data Mining.
- [15] 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.
- [16] M. Zaki, "SPADE: An efficient algorithm for mining frequent sequences," Kluwer Academic Publisher's Machine Learning, vol. 42, pp. 31-60, 2001.
- [17] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.C. Hsu, "Free Span: Frequent pattern projected sequential pattern mining," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '00), pp. 355-359, Aug. 2000.
- [18] KJ. 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.
- [19] 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.
- [20] 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.
- [21] J. Pei, J. Han, and R. Mao, "CLOSET: An efficient algorithm for mining frequent closed itemsets," Proceedings of ACM DMKD pp. 21-30, May 2000.
- [22] M. Zaki and C. Hsiao, "CHARM: An efficient algorithm for closed itemset mining," Proceedings of SIAM's SDM '02, pp. 457-473, Apr. 2002.
- [23] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Yongkang Wang, Helen Pinto, Qiming Chen Umeshwar Dayal, Mei-Chun Hsu, "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan approach", IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 10, October 2004
- [24] Thomas. Rincy. N, Yogadhar Pandey, "Performance Evaluation on the State of the art of Sequential Pattern Mining Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 65– No.14, March 2013.
- [25] Mahdi Esmaeili and Fazekas Gabor, "Finding Sequential Patterns from Large Sequence Data", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, January 2010 ISSN (Online): 1694-0784 ISSN (Print): 1694-0814.
- [26] Niti Desai, Amit Ganatra, "Sequential Pattern Mining Methods: A Snap Shot", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 10, Issue 4 (Mar. - Apr. 2013), PP 12-20.
- [27] Tzung-Pei, Hong, Ching-Yao Wang and Shian-Shyong Tseng, "An Incremental Mining Algorithm for Maintaining Sequential Patterns Using Pre-large Sequences," Journal Expert Systems with Applications, Vol. 38, Issue 6, p p. 7051-7058, 2011.
- [28] Jen-Wei Huang, Chi-Yao Tseng, Jian-Chih Ou, Ming-Syan Chen, "A General Model for Sequential Pattern Mining with a Progressive Database," IEEE Transactions on Knowledge and Data Engineering, vol. 20, No. 9, pp. 1153-1167, 2008.
- [29] Jiaxin Liu, "The design of storage structure for sequence in incremental sequential patterns mining," Networked Computing and Advanced Information Management (NCM), pp. 330 - 334, 2010.
- [30] Philippe Fournier, Viger, Roger Nkambou and Vincent Shin-Mutseng, "Rule Growth: Mining Sequential Rules Common to Several Sequences by Pattern-Growth," Symposium on Applied Computing, pp. 951-960, 2011.
- [31] Jiaxin Liu, "The design of frequent sequence tree in incremental mining of sequential patterns," Software Engineering and Service Science (ICSESS), pp. 679-682, 2012.