IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 4, May 2010
ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

An Efficient System Based On Closed Sequential Patterns for Web Recommendations

Utpala Niranjan[1], Dr.R.B.V. Subramanyam[2], Dr.V.Khanaa[3]
[1] Department of Information Technology, Rao & Naidu Engineering College, Ongole, Andhra Pradesh, 523 001, India,

[2] Department of Computer Science and Engineering, National Institute of Technology, Warangal, Andhra Pradesh 506 004, India,

[3] Department of Information Technology, Bharath University, Chennai, Tamilnadu 600 020, India

Abstract

Sequential pattern mining, since its introduction has received considerable attention among the researchers with broad applications. The sequential pattern algorithms generally face problems when mining long sequential patterns or while using very low support threshold. One possible solution of such problems is by mining the closed sequential patterns, which is a condensed representation of sequential patterns. Recently, several researchers have utilized the sequential pattern discovery for designing a web recommendation system, which provides personalized recommendations of web access sequences for users. This paper describes the design of a web recommendation system for providing recommendations to a user's web access sequence. The proposed system is mainly based on mining closed sequential web access patterns. Initially, the PrefixSpan algorithm is employed on the preprocessed web server log data for mining sequential web access patterns. Subsequently, with the aid of post-pruning strategy, the closed sequential web access patterns are discovered from the complete set of sequential web access patterns. Then, a pattern tree, a compact representation of closed sequential patterns, is constructed from the discovered closed sequential web access patterns. The Patricia trie based data structure is used in the construction of the pattern tree. For a given user's web access sequence, the proposed system provides recommendations on the basis of the constructed pattern tree. The experimentation of the proposed system is performed using synthetic dataset and the performance of the proposed recommendation system is evaluated with precision, applicability and hit ratio.

Keywords: Closed Sequential pattern mining, Web Personalization, Web server log data, Prefix span, Pattern tree, Patricia trie.

1. Introduction

The development of data mining techniques has been centralized on discovering hidden data in an efficient way that is beneficial for corporate decision-makers [1, 2]. Sequential pattern mining is an important subject of data mining which is extensively applied in several areas [3]. In general, sequential pattern mining is defined as determining the complete set of frequent subsequences in a set of sequences [4, 13]. Sequential pattern is a sequence of itemsets that frequently appear in a particular order, such that, all items in the same itemset are expected to have the same transaction time value or within a time gap. Each sequence relates to a temporally ordered list of events, where each event is considered as a collection of items/itemset occurring at the same time [5]. Generally, all the transactions of a customer are collectively considered as a sequence, known as customer-sequence, where each transaction is denoted as an item set in that sequence and all the transactions are listed in a particular order in connection with the transaction-time [6].

Normally, sequential pattern mining algorithms determine all frequent sequential patterns, defended by a minimum support from a database of sequences. This mining algorithm has an outcome of the following two problems. (1) First, sequential pattern mining generates large number of candidate patterns in an exponential curve regularly, which is unavoidable when the database comprises of long frequent sequential patterns. For instance, let us assume that the database holds a frequent sequence $i_1, i_2, \ldots, i_k$, k=20, it will generate $(2 \wedge 20) - 1$ frequent subsequences which are basically redundant patterns. (2) Second, establishing minimum support is also a challenging task for knowledge works: The smaller value of minimum support may lead to generate larger amount of candidate sequential patterns, while a larger value may result in no answer [7]. A closed sequential pattern mining is developed to address the challenges in mining all frequent sequential patterns like the large number of mined patterns, low support threshold and the long running time [8, 9].

A closed sequential pattern is said to be a specialized sequential pattern which has no super-sequence with the same occurrence frequency [10]. There exist two

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

approaches for mining closed frequent patterns: (1) determine the final closed pattern set (2) determine a closed pattern candidate set and perform post-pruning on it (for the extreme cases, perform on-the-fly checking, to be exact., for each newly determined pattern verify the previous patterns to notice if it is closed with respect to the discovered patterns). It seems desirable to employ the first approach owing to the reason that the second needs memory space to store the discovered patterns and perform post-pruning; but, when the patterns turn out to be more complex, it is hard to assure that each generated pattern is closed without examining the formerly discovered patterns [11]. The mining of closed sequential pattern have fascinated most of the researchers for its potential of employing compact results to maintain the equivalent expressive power as traditional mining [12].

In recent years there has been an overwhelming attention in employing sequential mining techniques to construct personalization systems. In order to obtain a compact result, usage of closed sequential pattern mining is more useful as it retains all the potential information. 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 [14]. 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 [15]. Web Recommender system is a kind of personalized web application which supplies considerable user value by personalizing a number of sites on the Web [16].

Web recommender systems have been drawing attention to a greater extent as an appropriate approach to counter information overload and aid the users of the Web information space to find what they require in a faster way [17]. Web recommendation systems [18] are useful in instructing the users to the target pages. On the other hand, employing recommendation systems to available Web sites needs considerable work and knowledge [19]. There exist many methods to build such systems. Among these methods, the building of web recommendation system by web access log is an accepted method [20], [21]. In this paper, we have designed a web recommendation system based on closed sequential pattern. The proposed system can be used for generating a personalized Web experience for the user. The input to the proposed system is the web server log data. The web server log data describes the users visiting behavior and generally consists of the following fields: IP address, access time, HTTP request method used, URL of the referring page and browser name. It is difficult to mine frequent sequential web access patterns from the raw web server log data directly. So, the web server log data is to be preprocessed and then sequential web access patterns are mined from it. Then, closed sequential web access patterns are discovered from the mined sequential web access patterns. Subsequently, a pattern tree is constructed using the mined closed sequential web access patterns. For a current user access sequence, web recommendations can be given to the user by matching the constructed pattern tree.

The rest of the paper is organized as follows: Section 2 presents a brief review of the researches related to the proposed system. Section 3 details the proposed web recommendation system. The experimental results and performance evaluation of the proposed system are given in Section 4 and finally the conclusions are summed up in Section 5.

2. Review of Related Research

Numerous researches are available in the literature for web recommendation system using sequential pattern mining. It looks more preferable if we design web recommendation system based on closed sequential patterns for its compact representation. Here, we present some of the researches related with closed sequential pattern mining along with web recommendation system based on sequential pattern mining.

Zhou. B *et al.* [26] have proposed an intelligent Web recommender system identified as SWARS (sequential Web access-based recommender system) that employs sequential access pattern mining. In the proposed system, CS-mine, an efficient sequential pattern mining algorithm was made use of to recognize frequent sequential Web access patterns. The access patterns were then stored in a compact tree structure (Pattern-tree) which was then employed for matching and generating Web links for recommendations. The performance of the proposed system was assessed on the basis of precision, satisfaction and applicability. An efficient sequential access pattern mining algorithm, called CSB-mine (Conditional Sequence Base mining algorithm) was presented by Baoyao Zhou *et al.* [27]. The presented CSB-mine algorithm was on the basis of conditional sequence bases of each frequent event which removes the need for constructing WAP-trees. This enhanced the efficiency of the mining process considerably in comparison to WAP-tree based mining algorithms, particularly when the value of support threshold becomes smaller and the database size gets larger. The proposed CSB-mine algorithm and its performance were discussed. They have also described a sequential access-based web recommender system that has included the CSB-mine algorithm for web recommendations.

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

Cui Wei *et al.* [28] have presented a hybrid web personalization system which was on the basis of clustering and contiguous sequential patterns. Their system clustered log files to find out the basic architecture of websites, and for each cluster, they employed contiguous sequential pattern mining to optimize the topologies of websites further. They have presented two evaluating parameters to test the performance of our system. Zhenglu Yang *et al.* [13] have presented an efficient sequential mining algorithm (LAPIN_WEB: LAst Position INduction for WEB log), which is an extension of previous LAPIN algorithm to extract user access patterns from traversal path in Web logs. Web log mining system comprises of data preprocessing, sequential pattern mining and visualization. The experimental results and performance studies established that LAPIN_WEB was efficient and outplayed familiar PrefixSpan by up to an order of magnitude on real Web log datasets. Furthermore, they also implemented a visualization tool to aid interpret mining results and also forecast users' future requests.

Kuo-Yu Huang *et al.* [29] have presented an approach which extends a frequent sequence with closed itemsets instead of single items. The closed sequential patterns were made up of only closed itemsets. Therefore, needless item extensions which generate non-closed sequential patterns were prevented. Experimental results proved that the proposed approach was two orders of magnitude faster than the existing related works with a reasonable memory cost. An efficient algorithm, known as TSP (top-k closed sequential patterns), was developed by Petre Tzvetkov *et al.* [30] for mining closed patterns without min_support. Starting at (absolute) min_support=1, the algorithm used the length constraint and the properties of top-k closed sequential patterns to execute dynamic support raising and projected database pruning. The performance study illustrated that TSP has high performance. It outplayed the efficient closed sequential pattern-mining algorithm, CloSpan, even when the latter was processed with the best tuned min_support threshold.

Takeaki Uno *et al*. [31] have presented an efficient algorithm LCM (Linear time Closed pattern Miner) for mining frequent closed patterns from large transaction databases. The major contribution towards the theoretical part was the proposed prefix-preserving closure extension of closed patterns, which allowed them to look for all frequent closed patterns in a depth-first manner, in linear time for the number of frequent closed patterns. Their algorithms do not require any storage space for the earlier obtained patterns, while the existing algorithms require it. Performance analysis of LCM with straightforward algorithms illustrated the positive aspects of prefix-preserving closure extension. Nancy P. Lin *et al.* [7] have presented an algorithm for mining closed frequent sequences, a scalable, condensed and lossless structure of complete frequent sequences that was mined from a sequence database. The algorithm, FMCSP (Fast Mining of Closed Sequential Patterns), has employed a number of optimization methods, namely equivalence class, to lessen the needs of searching space and run time. Specifically, one of the main problems in this type of algorithms was the redundant generation of the closed sequences; therefore, they presented an efficient and memory saving methods, diverse from existing works, that does not require the complete set of closed sequences to be residing in the memory.

3. A Proposed Web Recommendation System Based On Closed Sequential Patterns

Web Personalization is an application of data mining and machine learning techniques to build models of user behavior. It can be useful to the task of predicting user needs and adapting future interactions with the main aim of improved user satisfaction [22]. A unique and important class of personalized Web applications is represented by Web recommendation systems. It highlights on the user-dependent filtering and selection of relevant information. Several approaches like Content-Based Filtering, Clustering Based Approaches, Graph Theoretic Approaches, Association and Sequence Rule Based Approaches are available in the literature for designing web recommendation system [22]. Here, the proposed web recommendation system is designed based on the closed sequential patterns. The proposed system is used for generating a personalized Web experience for a user. The block diagram of the proposed web recommendation system based on closed sequential patterns is given in Fig 1. The important steps for generating recommendations to the user is as follows,

- Preprocessing
- Mining of sequential web access patterns
- Mining of closed sequential web access patterns
- Construction of Pattern tree
- Generation of recommendations for user web access sequences

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

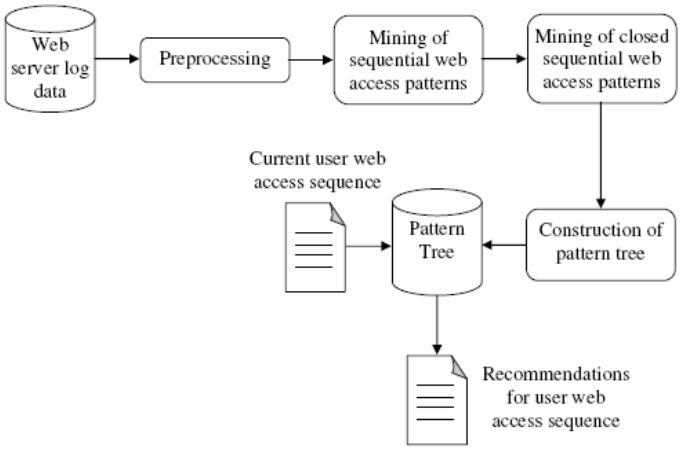

Fig 1: Block diagram of the proposed web recommendation system

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

3.1 Preprocessing

The purpose of data preprocessing is to extract useful data from raw web log and then transform these data in to the form necessary for pattern discovery. Due to large amount of irrelevant information in the web log, the original log cannot be directly used in the web log mining procedure, hence in data preprocessing phase, raw Web logs need to be cleaned, analyzed and converted for further step. A Web log is a file to which the Web server writes information each time a user requests a resource from that particular site. Most logs use the format of the common log format. Each entry in the log file consists of a sequence of fields relating to a single HTTP transaction with the various fields.

The input for the proposed web recommendation system is a web server log data and it comprises IP address, access time, HTTP request method used, URL of the referring page and browser name (for an instance, *Web server log file: 192.162.26.12 [12/Oct/2009:11:17:55] "GET / HTTP/1.1" "http://en.wikipedia.org/wiki/Association_rule_learning" Mozilla/5.0 Windows xp*). It is difficult for these web server log data to be directly used to mine the desired sequential pattern mining process. So, due to that phenomenon, the following preprocessing techniques need to be used in the raw web server log data.

(1) User Identification: User identification means identifying each user accessing web page, whose goal is to mine every user's access characteristic. Users may be tracked based on the IP address of the computer requesting the page and user sessions. A new IP address is used to identifying the new user and at the same time, the user session must be within certain time limits.
(2) Mapping: For every identified user, the visiting web pages are arranged into row-wise in such a way that it forms the sequential database D.

3.2 Mining of Sequential web access Pattern

Sequential web access patterns are mined from the sequential database D, which is a set of 2-tuples (sid, α), where sid is a user-id and α is a sequence of web pages accessed by users. The problem of mining sequential pattern is defined as: Let $W = \{z_1, \ldots, z_n\}$ be a set of web access events. A sequence $\alpha = \langle Z_1 \ldots Z_l \rangle$ is an ordered list of web access events. A web access event $Z_i (1 \le i \le l)$ in a sequence may have a special attribute like, timestamp denoted as $Z_i.t$, which registers the time when the web access event Z_i was executed. As a notational convention, for a sequence $\alpha = \langle Z_1 \cdots Z_l \rangle, Z_i.t < Z_j.t$ for $1 \le i < j \le l$.. The number of web access event in a sequence is called the length of the sequence. A sequence α with length l is called an l-sequence, denoted as $len(\alpha) = l$. A sequence $\alpha = \langle Z_1 \ldots Z_n \rangle$ is called a subsequence of another sequence $\beta = \langle Y_1 \cdots Y_m \rangle$, $(n \le m)$ and β a super-sequence of α, denoted as $\alpha \subseteq \beta$, if there exist integers $1 \le i_1 < \cdots < i_n \le m$ such that $Z_1 \subseteq Y_{i_1}, \cdots, Z_n \subseteq Y_{i_n}$. A tuple (sid, α) in a sequence database S is said to contain a sequence γ, if γ is a subsequence of α. The number of tuples in a sequence database S containing sequence γ is called the support of γ, denoted as $\sup(\gamma)$. Given a positive integer $\min\_\sup$ as the support threshold, a sequence γ is a sequential pattern in sequence database S if $\sup(\gamma) \ge \min\_\sup$ [23].

3.2.1 PrefixSpan algorithm

The proposed recommendation system utilizes PrefixSpan, a well-known pattern-growth algorithm, for mining the complete set of sequential web access pattern from the sequential database D. The main advantage of PrefixSpan is the use of projected databases. An *a*-projected sequence database is the set of subsequences in the sequence database that are suffixes of the sequences that have the prefix *a*. In every step, the algorithm checks for the frequent sequences with prefix a, in the correspondent projected database [24]. The algorithmic description of PrefixSpan algorithm is shown in Fig 2.

Input : A sequence database S, and the minimum support threshold min_sup

Output : The complete set of sequential patterns

Method: Call *PrefixSpan*$(\langle\ \rangle, 0, S)$.

Subroutine: *PrefixSpan* $(\alpha, l, S|_\alpha)$

Parameters: α : a sequential pattern; l : the length of α ; $S|_\alpha$: the α-projected database, if $\alpha \ne \langle\ \rangle$; otherwise, the sequence database S.

Method:

1. Scan $S|_\alpha$ once, find the set of frequent items b such that
   a) b can be assembled to the last element of α to form a sequential pattern; or
   b) $\langle b \rangle$ can be appended to α to form a sequential pattern.
2. For each frequent item b, append it to α to form a sequential pattern α', and output α';
3. For each α', construct α'-projected database $S|_{\alpha'}$, and call *PrefixSpan* $(\alpha', l+1, S|_{\alpha'})$.

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

Fig 2: Pseudo Code of the PrefixSpan algorithm

3.3 Mining of Closed Sequential Web Access Patterns

This sub-section details the mining of closed sequential web access patterns from the complete set of sequential web access patterns (FS) resulted from the previous sub-section. The closed patterns are mined with the aid of the post-pruning strategy. The definition of closed sequential patterns is as follows:

Definition: The set of closed frequent sequential pattern is defined as follows.

$$CS = \{\alpha \mid \alpha \in \text{FS and} \nexists \beta \in \text{FS such that} \alpha \subseteq \beta \text{ and support}(\alpha) = \text{support}(\beta)\}$$

The recommendation system used with the closed sequential pattern mining has the same expressive power as compared with the regular sequential pattern mining. Additionally, the closed sequential patterns have the capability to provide the compact result set (reducing the generation of the number of sequential patterns) [25].

First, each sequential web access pattern in the pattern set FS_i is compared with the other patterns in the set, for instance $_{FS_j}$. The patterns $_{FS_i}$ is removed from the pattern set FS, if and only if, (1) the support of both web access patterns $_{FS_i}$ and FS_j should be same and (2) FS_i must be a subset of $_{FS_j}$ ($FS_i \subseteq FS_j$). At the end, we obtain the closed pattern set CS ($CS \subseteq FS$) from the pattern set FS.

Example: Consider a sequential database of web access sequence D in Table 1. The complete set of sequential patterns with $\min\_\sup = 2$ are FS = {a:4, aa:2, ab:4, abb:2, abc:4, ac:4, b:4, bb:2, bc:4, c:4, ca:3, cab:2, cabc:2, cac:2, cb:3, cbc:2, cc:2}. The set FS consists of 17 sequences. From these 17 frequent sequential patterns, the closed sequential patterns discovered are: $_{CS}$ = {aa:2, abb:2, abc:4, ca:3, cabc:2, cb:3}. From the above example, it is clear that closed sequential pattern set CS is a compact list of frequent sequential pattern set FS.

User id	Web access sequence
1	c a a b c
2	a b c b
3	c a b c
4	a b b c a

Table 1: Sample sequential database of web access sequences

3.4 Construction of Pattern Tree

The construction of a pattern tree from the mined closed sequential web access patterns is described in this sub-section. The constructed pattern tree is used in making the recommendation for a user's web access sequence. The constructed pattern tree is based on the Patricia trie (Radix tree or crit bit tree) [32] data structure. In general, the Patricia trie is used to store a set of strings but in regular trie, single character is stored in each node. By using Patricia trie, the tree can be even more compacted when compare with the regular trie. The procedure for constructing the pattern tree in the proposed system is as follows:

1) Create an empty root node.
2) Insert the most sub pattern in the closed pattern set CS into a node next to the root node.
3) Insert the postfixes of pattern into child node if the current pattern to be inserted is a super pattern of inserted patterns.
4) Otherwise, current pattern is inserted into the node next to the root node.
5) Step 3 and step 4 is repeated for every pattern in the closed pattern set $_{CS}$.

.

Example: The closed web access sequential pattern CS = {aa: 2, abb: 2, abc: 4, ca: 3, cabc: 2, cb: 3}. Initially, we create an empty root node. Then, web access patterns <aa: 2>, <abb: 2>, <abc: 4> and <ca: 3> are inserted next into the root node based on the procedure depicted above. Web access pattern <cabc: 2> is a super pattern of <ca> so that the postfix of <ca> i.e., <bc> is inserted into a child node of <ca>. Moreover, <cb> is also inserted next to the root node. The pattern tree for the chosen example is shown in Fig 3.

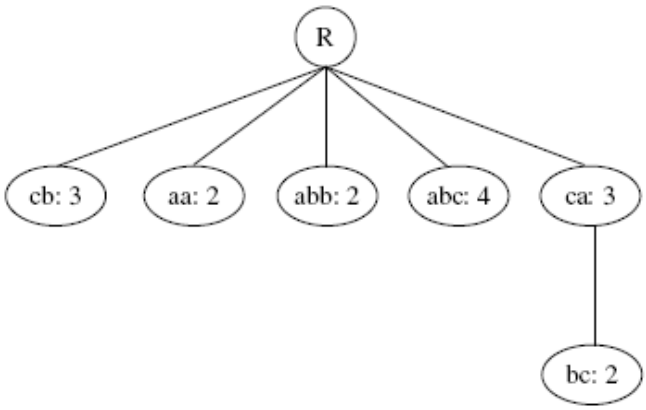

Fig 3: The Pattern-tree constructed from the closed sequential web access patterns

3.5 Generation of Recommendations for user access sequences

The constructed pattern tree is used for matching and generating web links for recommendations for user access

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

sequences. Once the pattern tree is constructed, instead of using the sequential database of web access sequence D, we can use the constructed pattern tree for the generation of recommendations. The recommendations are retrieved for a given user's web access sequence S, length of the user web access sequence S must satisfy the thresholds (*minlen* and *maxlen*). For a user's web access sequence, initially, the sequences with the presence of the user access sequence are identified from the constructed pattern tree. Then, the postfixes of those patterns are provided as recommendation to the user's web access sequence. For a pattern without postfixex, it's child node are provided as recommendations.

For instance, the proposed system provides recommendations R for user web access sequences <ab> and <ca> as follows:

Case 1*:* The constructed pattern tree is searched for sequential patterns with the presence of User's web access sequence <ab>. In the pattern tree (shown in figure 3), next to the root node has the sequences <abb: 2> and <abc: 4> that are subset of given user web access sequence <ab>. So the postfixes of these sequential patterns i.e., <b> and <c> are given as recommendations to the user (R = b: 2 and c: 4).

Case 2: For user web access sequence <ca>, the pattern tree (shown in figure 3) is searched for pattern with the presence of <ca>. As the pattern <ca> does not have any postfixes, it's child node i.e., <b> is provided as recommendations for the user (R = b: 2).

4. 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 in Java (jdk 1.6). 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.

4.1 Evaluation Measures

By evaluating the proposed system, we have used three measures such as precision, applicability and hit ratio. The formal definition of these three measures is given as,

$$Precision = \frac{R^+}{R^+ + R^-}$$

Where, R^+ → Number of correct recommendations.

R^- → Number of incorrect recommendations.

Definition: Let $S = s_1\ s_2\ \cdots\ s_j\ s_{j+1}\ \cdots\ s_n$ be a web access sequence of test dataset. The recommendation $R = \{r_1, r_2, \cdots, r_k\}$ is generated by using the constructed pattern tree for the subsequence $S_{sub} = s_1\ s_2\ \cdots\ s_j$ $(\text{minlen} \le j \le \text{maxlen})$. The recommendation R is said to be correct, if it contains s_{j+1} $(s_{j+1} \in R)$. Otherwise, R is said to be incorrect recommendation.

$$Applicability = \frac{R^+ + R^-}{|R|}$$

Where, $|R|$ → Total number of given requests.

$$Hit\ ratio = Precision \times Applicability = \frac{R^+}{|R|}$$

4.2 Experimental Results

The experimental results of the proposed recommendation system are presented in this section. We measure the precision, applicability and the hit ratio for different support thresholds and the results are plotted in graph as shown in figure 4. The comparison of sequential pattern mining (SPM) with closed sequential pattern mining (Closed SPM) in terms of generated web access patterns is given in Table 2 and the corresponding graph is shown in fig 5.

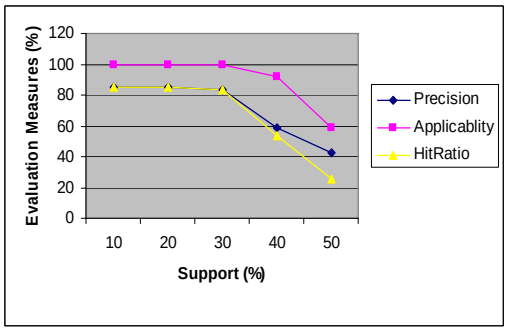

Fig 4: Evaluation measures vs. Support threshold

Support	SPM	Closed SPM
10	6203	4760
20	1125	1027

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

30	559	470
40	218	201
50	92	85

Table 2: Number of patterns generated for SPM and closed SPM

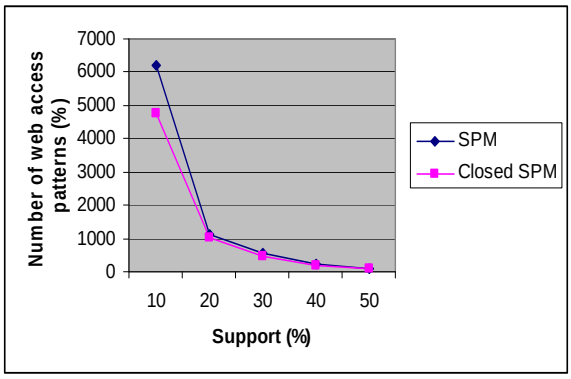

Fig 5: Comparison graph of SPM and Closed SPM.

5. Conclusion

In this paper, we have devised a web recommendation system based on closed sequential pattern mining. In the proposed system, we have employed PrefixSpan (pattern growth algorithm) for mining the sequential web access patterns from the preprocessed web server log data and the closed sequential web access patterns are obtained from the mined sequential web access pattern. Then, Patricia trie based data structure is used to build the pattern tree, where mined closed sequential web access patterns are stored. The target web link is given to the new user by matching the pattern tree with the new user's current web access sequence. The proposed web recommendation system is validated with the synthetic dataset and the experimental results showed that the proposed system outperformed with good precision, applicability and hit ratio.

References

[1] M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley & Sons Inc., New York, 2002.

[2] M.S. Chen, J. Han and P.S. Yu, "Data mining: an overview from a database perspective", IEEE Transactions on Knowledge and Data Engineering, vol. 8, pp. 866-883, 1996.

[3] Sizu Hou, Xianfei Zhang, "Alarms Association Rules Based on Sequential Pattern Mining Algorithm," In proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 2, pp.556-560, Shandong, 2008.

[4] R. Agrawal and R. Srikant, "Mining Sequential Patterns", In Proceedings of the 11th International Conference on Data Engineering, pp. 3-14, Taipei, Taiwan, 1995.

[5] Salvatore Orlando, Raffaele Perego and Claudio Silvestri, "A new algorithm for gap constrained sequence mining", In Proceedings of the ACM Symposium on Applied Computing, pp. 540 – 547, Nicosia, Cyprus, 2004.

[6] Qiankun Zhao and Sourav S. Bhowmick, "Sequential Pattern Mining: A Survey", Technical Report, CAIS, Nanyang Technological University, Singapore, no.118, 2003.

[7] Nancy P. Lin, Wei-Hua Hao, Hung-Jen chen, Hao-En Chueh, Chung-I chang, "Fast Mining of Closed Sequential Patterns", WSEAS Transactions on Computers, vol. 7, no. 4, pp. 133-139, 2008.

[8] Claudia Antunes and Arlindo L. Oliveira, "Sequential Pattern Mining with Approximated Constraints", in proceedings of the International Conference on Applied Computing, pp. 131-138, 2004.

[9] Congnan Luo, Soon M. Chung, "Efficient Mining of Maximal Sequential Patterns Using Multiple Samples", in proceedings of International Conference on Signal Processing Systems, pp.788-792, 2005.

[10] Shengnan Cong, Jiawei Han and David Padua, "Parallel Mining Of Closed Sequential Patterns", in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 562 – 567, Chicago, Illinois, USA, 2005.

[11] Yan. X, Han. J and Afshar. R, "CloSpan: mining closed sequential patterns in large datasets", In Proceedings of the 3rd SIAM International Conference on data mining, pp. 166-177, San Francisco, CA, May 2003.

[12] Ming-Yen Lin, Sue-Chen Hsueh and Chia-Wen Chang, "Mining Closed Sequential Patterns with Time Constraints", Journal of Information Science and Engineering, vol. 24, pp. 33-46, 2008.

[13] Zhenglu Yang, Yitong Wang and Masaru Kitsuregawa, "An Effective System for Mining Web Log", LNCS, vol. 3841, pp.40-52, 2006.

[14] Magdalini Eirinaki and Michalis Vazirgiannis, "Web Mining for Web Personalization", ACM Transactions on Internet Technology (TOIT), vol. 3, no.1, pp. 1 - 27 , 2003.

[15] Mulvenna. M. D, Anand. S. S and Buchner. A. G, "Personalization on the Net using Web Mining", . Communications of the ACM, vol. 43, no. 8, pp. 123–125, August 2000.

[16] J. Ben Schafer, Joseph A. Konstan and John T. Riedl, "Recommender Systems for the Web", In Visualizing the Semantic Web, Springer, pp.102-123, 2006.

[17] Olfa Nasraoui, Zhiyong Zhang, and Esin Saka, "Web Recommender System Implementations in Multiple Flavors: Fast and (Care) Free for All", in Proceedings of SIGIR Open Source Information Retrieval Worskhop, Seattle, WA, July 2006.

[18] J. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-Commerce", in Proceedings of ACM Conference on Electronic Commerce (EC-99), pp.158-166, 1999.

ISSN (Online): 1694-0784
ISSN (Print): 1694-0814

[19] Hiroshi Ishikawa, Manabu Ohta, Shohei Yokoyama, Junya Nakayama, and Kaoru Katayama, "On the Effectiveness of Web Usage Mining for Page Recommendation and Restructuring", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol. 2593, pp. 253-267, 2009.
[20] Osmar R. Zaiane, Jia Li, and Robert Hayward, "Mission-Based Navigational Behaviour Modeling for Web Recommender Systems", in proceedings of the 6th International workshop on knowledge discovery, Seattle, WA, USA, August 22-25, 2004.
[21] Mehdi Adda, Petko Valtchev, Rokia Missaoui and Chabane Djeraba, "Toward Recommendation Based on Ontology Powered Web-Usage Mining", Internet Computing, vol. 11, no.4, pp. 45-52, 2007.
[22] Sarabjot Singh Anand and Bamshad Mobasher, "Intelligent Techniques for Web Personalization", Lecture Notes in Computer Science, Vol. 3169, Springer, 2005.
[23] Jian Pei, Jiawei Han and Wei Wang, "Mining sequential patterns with constraints in large databases", in Proceedings of the eleventh international conference on Information and knowledge management, McLean, Virginia, USA, pp. 18 - 25, 2002.
[24] Claudia Antunes and Arlindo L. Oliveira, "Generalization of Pattern-growth Methods for Sequential Pattern Mining with Gap Constraints", Lecture Notes in Computer Science, Vol. 2734, pp.239-251, 2003.
[25] Jianyong Wang and Jiawei Han, "BIDE: Efficient Mining of Frequent Closed Sequences", in Proceedings of 20th International Conference on Data Engineering, pp. 79- 90, 2004.
[26] Zhou. B,Hui. S.C and Chang. K, "An intelligent recommender system using sequential Web access patterns", IEEE Conference on Cybernetics and Intelligent Systems, vol. 1, pp. 393 - 398, December 2004.
[27] Baoyao Zhou, Siu Cheung Hui and Alvis Cheuk Ming Fong, "Efficient sequential access pattern mining for web recommendations", International Journal of Knowledge-based and Intelligent Engineering Systems, vol. 10 , no. 2, pp. 155-168, April 2006.
[28] Cui Wei, Wu Sen, Zhang Yuan and Chen Lian-Chang, "Algorithm of mining sequential patterns for web personalization services", ACM SIGMIS Database, vol. 40 , no. 2, pp. 57-66,May 2009.
[29] Kuo-Yu Huang , Chia-Hui Chang , Jiun-Hung Tung and Cheng-Tao Ho, "COBRA: Closed Sequential Pattern Mining Using Bi-phase Reduction Approach", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol. 4081, pp. 280-291, 2006.
[30] Petre Tzvetkov, Xifeng Yan and Jiawei Han, "TSP: Mining top-k closed sequential patterns", Knowledge and Information Systems, Vol. 7, No. 4, pp. 438-457, May 2005.
[31] Takeaki Uno, Tatsuya Asai , Yuzo Uchida and Hiroki Arimura, "An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, vol. 3245, pp. 16-31, 2004.
[32] Knizhnik, Konstantin. "Patricia Tries: A Better Index for Prefix Searches", Dr. Dobb's Journal, June, 2008.

First UTPALA NIRANJAN received M.Sc., and M.Tech., degrees from Department of Computer Science and Information Technology from Sri Venkateswara University, Tirupati and Bharath University, Chennai in 1999 and 2005 respectively. He is working in Rao & Naidu Engineering College, Ongole. He is currently pursuing the Ph.D. degree from Bharath University, Chennai, working closely with Dr.R.B.V Subramanyam, NIT Warangal and Dr.V.Khanaa, Bharath University Chennai.

Second Dr.R.B.V..SUBRAMANYAM received the M.Tech., and Ph.D degrees from IIT Kharagpur He is working in NIT Warangal, Andhra Pradesh. He works in the field of Data mining. He is one of the reviewer for IEEE Trans. On Fuzzy Systems and also for Journal of Information Knowledge Management. Membership in Professional bodies:1.Member, IEEE; 2. Associated Member, The Institution of Engineers (INDIA).

Third Dr.V.KHANA is working as Dean – Information Bharath University, Chennai.