International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Bidirectional Search on Cyclic Behavioral Patterns for Web Sequential Mining

A. P. Selva Prabhu[1], Dr. T. Ravi Chandran[2]

[1]Ph.D Research Scholar, Department of Computer Science, Research and Development centre, Bharathiar University Coimbatore and Assistant Professor, Bharathiar University Arts and Science College Sivagiri, Erode, Tamil Nadu, India

[2]Dean and Hod, Electronics and Communication Engineering, S.N.S. College of Technology,Coimbatore , Tamil Nadu, India

Abstract: ***Web sequential pattern mining is significant for evaluating the access behavior of web users. Many research works have been developed for web sequential patterns mining. But, conventional web sequential patterns mining techniques takes more time for extracting frequent web pages. Besides, the performance of existing web sequential patterns mining was not efficient. In order to solve these limitations, Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method is proposed. The BES-CBPM method is designed with objective of increasing the mining efficiency of web sequential patterns with minimum time and space complexity. The BES-CBPM method used Bidirectional Expansion Search algorithm for extracting the similar user's interest web pages from weblog database. The Expansion Search algorithm developed in BES-CBPM method simultaneously carried outs two search with aim of mining the similar user's interest web pages. This process resulting in reduced time and space complexity for web sequential patterns mining. In addition, BES-CBPM method used cyclic behavioral search based pattern mining with aiming at enhancing the mining efficiency. The cyclic behavioral search based pattern mining in BES-CBPM method reiterates the bidirectional expansion search for a certain period of time in order to efficiently improve mining performance of web sequential patterns mining. The BES-CBPM method conducts the experimental works on metrics such as time complexity, space complexity, mining efficiency, true positive rate of mining and scalability. The experimental results demonstrate that the BES-CBPM method is able to improve the mining efficiency and also reduces the time complexity of web sequential pattern mining when compared to state of the art works.***

Keywords: Bidirectional Expansion Search, Cyclic Behavioral Search, Mining, Web Sequential Pattern, Web Users, Weblog Database

1. Introduction

World Wide Web (WWW) contains massive amount of web pages and links that provides enormous information for internet users whom can be accessed. Each visitor of a web site leaves a trace in a weblog of the pages that he or she browsed. Sequential Pattern Mining is widely used in data mining for analyzing web usage, purchase behavior, and text mining etc. A considerable application of sequential mining techniques is mining web log accesses patterns. Thus, sequential pattern mining is an essential to examine the access behavior of web users. Few researches have been designed for mining the web sequential patterns. For example, an automatic annotation approach was presented in [1] for mining the similar web pages in web databases. But, the amount of time taken for extracting web pattern was higher. An Ant-based clustering algorithm was designed in [2] for predicting the maximum visited pages by probable user from website and to predicting future visit of potential user in web server log. However, mining efficiency of web sequential pattern mining was not sufficient.

An improved approach of Gap-BIDE algorithm was developed in [3] for discovering closed sequential patterns and mining user access patterns from web log data. However, time complexity of web sequential pattern mining was more. A novel algorithm was intended in [4] to mine time-constrained sequential patterns of Smartphone application usage log collected from web and to enhance the mining efficiency. However, the effectiveness of this algorithm was not at required level. A web classification algorithm was presented in [5] with aid of fuzzy association rule mining to classify the web pages into diverse web categories based on the manner in which they present in user sessions. However, mining efficiency of web sequential pattern was poor. A decision tree algorithm was designed in [6] to extract log files and extract knowledge from web data stream and makes training rules p to find out dissimilar information related to log file. The decision tree algorithm improves the accuracy of mining web sequential patterns. However, the memory space taken for storing mined web patterns are higher.

A Temporal Relational Rule Mining Approach was developed in [7] to find out hidden information from their web log data with higher mining accuracy. But, the performance of mining was not efficient. The Generalized Sequential Pattern (GSP) algorithm was presented in [8] to extract frequent web sequential patterns through formation of a tree with minimum time. However, mining efficiency of web sequential patterns was remained unaddressed. A weighted support method was intended in [9] to efficiently extract weighted access patterns in a web log database. The weighted support method reduces the time and space complexity of mining the web sequential patterns. But, scalability rate was poor. A Node Linkage Depth-First Traversal (NLDFT) algorithm was designed in [10] for sequential pattern mining with minimum runtime and lessening memory requirements. However, true positive rate of sequential pattern mining was remained unsolved. In order to overcome the above mentioned existing issues, Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method is designed.

The major contribution of BES-CBPM method is organized as follows, To enhance the performance of web sequential pattern mining with higher mining efficiency, BES-CBPM method is introduced with application of bidirectional

expansion search and cyclic behavioral search based pattern mining. To efficiently minimize the time and space complexity of web sequential pattern mining with higher true positive rate, bidirectional expansion search is applied in BES-CBPM method. The bidirectional expansion search accomplishes two searches simultaneously in order to mine the frequent web pages. To achieve higher efficiency for extracting the similar user's interest web pages, cyclic behavioral search based pattern mining is used in BES-CBPM method. The cyclic behavioral search repeats the bidirectional expansion search for a certain period of time and thereby significantly mines missed similar web pages in weblog.

The rest of the paper is ordered as follows. Section 2 describes the process of Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method with neat diagram. In Section 3 the simulation environment is presented and the results are explained in section 4. Section 5 describes the related works. Finally, the conclusion of the paper is presented in section 6.

2. Bidirectional Expansion Search Based Cyclic Behavioral Pattern Mining Method

The Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method is developed with objective of increasing the mining efficiency of web sequential patterns and reducing time and space complexity. The BES-CBPM method used Bidirectional Expansion Search algorithm in order to reduce the time and space complexity involved in mining web sequential patterns. The Bidirectional Expansion Search algorithm simultaneously runs two searches for extorting similar user's interest web pages. This helps for BES-CBPM method to minimize time for mining the web sequential patterns in an effective manner. The Bidirectional Expansion Search extracts only a similar user's interest web pages from weblog through similarity measurement. This helps for BES-CBPM method to minimize the memory space taken for storing web sequential patterns in an efficient manner.

During the Bidirectional Expansion Search, some of frequent web pages in weblog database may not be extracted. Therefore, Cyclic Behavioral Search based Pattern mining is used in BES-CBPM method. The Cyclic Behavioral Search based Pattern mining set time period in order to reiterate the bidirectional expansion search for effectually mining the similar user's interest web pages. The reiteration of bidirectional expansion search helps for BES-CBPM method to extracts the missed frequent patterns during the search of bidirectional expansion. Therefore, BES-CBPM method attains the higher mining efficiency for web sequential patterns. The overall architecture diagram of Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method is shown in below Figure 1.

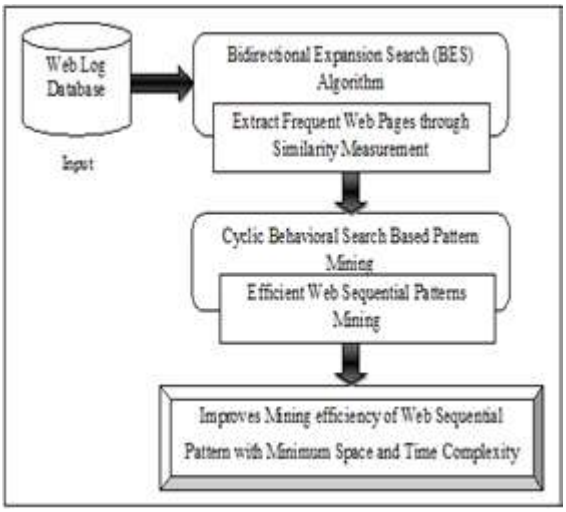

Figure 1: Architecture of Bidirectional Expansion Search based Cyclic Behavioral Pattern Mining

As demonstrated in Figure 1, BES-CBPM method initially takes weblog databases (i.e. Amazon Commerce reviews set dataset) as input. Then, BES-CBPM method applies Bidirectional Expansion Search algorithm with aiming at extracting the maximum visited web pages by web users in weblog with aid of similarity measurement. After that, BES-CBPM method used Cyclic Behavioral Search Based Pattern Mining for effectual web sequential patterns. As a result, BES-CBPM method enhances mining efficiency for web sequential patterns with minimum time and space complexity. The detailed process about BES-CBPM method is explained in forthcoming sub sections.

2.1 Bidirectional Expansion Search

The Bidirectional Expansion Search is used in BES-CBPM method with objective of mining the similar user's interest web sequential patterns from web log database and thereby achieving higher mining efficiency. The Bidirectional Expansion Search applied in BES-CBPM method searches the frequent web pages in two directions at the same time. In Bidirectional Expansion Search, one search is started from the initial state and the other search is begins backward from the goal. The Bidirectional Expansion Search is performed through expanding tree with branching factor and the distance from start to goal. The search of Bidirectional Expansion Search ends when both searches meet in the center point of graph. Bidirectional Expansion search designed in BES-CBPM method is a brute-force search algorithm that needs an explicit goal state. The process involved in Bidirectional Expansion Search for mining the sequential patterns from the weblog database is shown in below Figure 2.

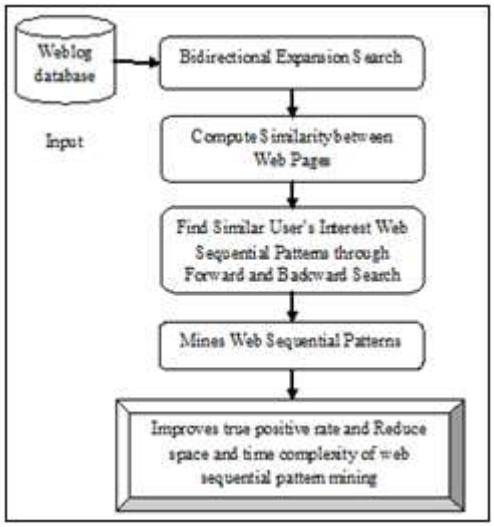

Figure 2: Process of Bidirectional Expansion Search for Web Sequential Pattern Mining

As shown in Figure 2, Bidirectional Expansion Search initially takes weblog database as input. After that, Bidirectional Expansion Search determines similarity among the web pages in weblog database with aid of jaccard similarity. Then, Bidirectional Expansion Search discovers the similar user's interest web patterns and then effectively extracts the web sequential patterns with help of forward and backward search. This helps for BES-CBPM method to increases mining efficiency of web sequential pattern with minimum space and time complexity.

Generally, search algorithm is used in data mining in order to mine information stored within some data structure. Therefore BES-CBPM method used Bidirectional Expansion Search (BES) algorithm with objective of mining the web sequential patterns from the web log database. The BES algorithm is a graph search algorithm for web sequential patterns mining which runs two simultaneous searches as follows,

1) Forward search (i.e. from initial state toward goal state)
2) Backward search (i.e. from goal state back toward initial state)

The BES algorithm replaces single search graph (which is likely to grow exponentially) with two smaller sub graphs in which one search initiates from initial vertex and other begins from goal vertex. The search of BES algorithm for finding similar user interest web pages ends when two search intersect. The BES algorithm is conducted through a heuristic estimate of remaining distance from source to goal and vice versa for finding shortest path possible to mine frequent web pages in weblog database. In BES algorithm, both initial and goal states are distinctive and completely defined. The following example diagram explains the processes of BES algorithm for web sequential pattern mining.

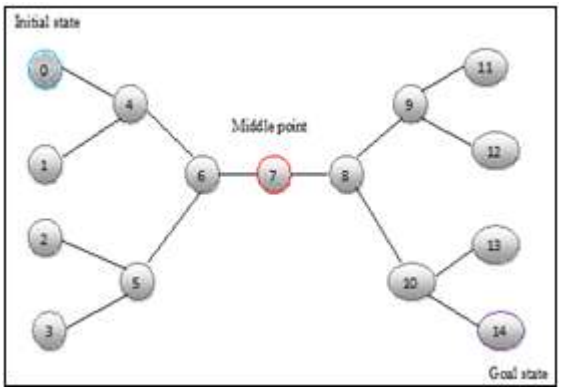

Figure 3: Graph Structure of Bidirectional Expansion Search for Web Sequential Pattern Mining

Figure 3 shows the graph structure of BES algorithm for extracting web sequential patterns in web log database. Each node in graph includes of different web pages browsed by different web users. Consider that we want to extract frequent web pages visited by web users from node 0 to node 14. The proposed BES algorithm carried outs two searches, one from node 0 and other from node 14. When both forward and backward search meet at node 7, BES algorithm efficiently finds and extracts similar user's interest web pages in weblog database and also search process of BES algorithm is terminated. Thus, BES algorithm significantly avoids unnecessary exploration of web sequential pattern mining. This helps for BES-CBPA method to minimize the amount of time taken for mining web sequential patterns and also to achieve higher mining efficiency.

During the forward search in BES algorithm, the search of web sequential patterns starts from initial state $x_{initial}$. Here, the current search state is represented as x. The next state of current state x is denoted as x'. For each node in graph $u \in U(x)$, the forward search for discovering similar user's interest web pages is performed using below mathematical expression,

$$x' = f(x, u) \quad (1)$$

From equation (1), x' denotes the next state of current state for finding frequent web pages whereas u represents forward search of edge. Similarly, the backward search of BES algorithm for identifying similar user's interest web pages is mathematically expressed as,

$$x' = f^{-1}(x', u^{-1}) \quad (2)$$

From equation (2), the interpretation of f^{-1} is simple to determine in terms of the state transition graph (i.e. reverse the direction of every edge in graph to identify web pattern). Here, u^{-1} indicates the reversed edge (i.e. represents a backward search). During the Bidirectional Expansion Search, the frequent web pages visited by web users are identified with assist of similarity measure. The BES algorithm used jaccard similarity for computing similarity between web pages in weblog database. The jaccard similarity between two web pages such as wp_i and wp_j is measured by using below mathematical formula,

$$Sim_{wp_i wp_j} = \frac{n_{wp_i wp_j}}{n_{wp_i} + n_{wp_j} - n_{wp_i wp_j}} \quad (3)$$

From equation (3) wp_i represent number of content features which appear in web page i whereas n_j denotes the number of content features which appear in web page j . Here, $n_{wp_i wp_j}$ indicates the number of content features that appear in both web pages. The Jaccard Similarity value is lies in the range of 0 to 1.0. Thus in weblog database, web pages which have jaccard similarity value between 0 and 1.0 is considered as similar users interest web pages. The remaining web pages in weblog database are considered as dissimilar webpage's. The algorithmic process of Bidirectional Expansion Search for mining web sequential patterns is shown in below,

// Bidirectional Expansion Search Based Web Sequential Pattern Mining Algorithm
Input: Amazon Commerce reviews set dataset
Output: Reduced Space And Time Complexity with higher true positive rate
Step 1: Begin
Step 2: Construct graph using web pages in weblog database
Step 3: For each web pages
Step 4: Compute similarity between web pages using (3)
Step 5: End for
Step 6: While (forward and backward search is meet each other at some point in the middle graph) **do**
Step 7: Identify and extracts similar users interest web pages through forward and Backward search
Step 8: Forward search for discovering similar users interest web pages is performed using (1)
Step 9: Backward search for discovering similar users interest web pages is performed using (2)
Step 10: End while
Step 11:End

Algorithm 1 Bidirectional Expansion Search Based Web Sequential Pattern Mining Algorithm

As shown in algorithm 1, the BES algorithm effectively extracts the frequent web pages visited by web users stored in weblog database with higher true positive rate. Initially, the BES algorithm takes dataset as input. After that, the BES algorithm builds graph by using web pages in dataset. Followed by, the BES algorithm evaluates the similarity between web pages in weblog database with help of jaccard similarity. Then, the BES algorithm find outs similar users interest web pages in weblog database by means of forward and backward search and subsequently extracts the web sequential patterns effectively. Both searches of BES algorithm terminated when they meet each other at some point in the middle of the graph.

The BES algorithm performs two simultaneous searches for significantly mining web sequential patterns which resulting in reduced time complexity. Besides, BES algorithm extracts only a frequent web pages browsed by web users which in turn helps for minimizing the memory space taken for storing the web sequential patterns in an effective manner. The BES algorithm guarantees optimal solutions for mining web sequential patterns from weblog database through performing two simultaneous searches (i.e. forward and backward search). This supports for BES-CBPM method to increase the mining efficiency. In order to further enhance the mining efficiency of similar users interest web pages, cyclic behavioral search based pattern mining is employed in BES-CBPM method which detailed explained in below sub section.

2.2 Cyclic Behavioral Search based Pattern mining

The cyclic behavioral search based pattern mining is used in BES-CBPM method in order to reiterate the bidirectional expansion search process. By reiteration of the bidirectional expansion search process using cyclic behaviours, BES-CBPM method mines the missed similar user's interest web patterns in weblog database efficiently. The cyclic behavioral search deals with the tendency of certain sequential patterns to repeat the bidirectional expansion search themselves periodically after definite time intervals. This concept is significantly helpful in predicting future browsing patterns. This helps for BES-CBPM method to achieve higher mining efficiency.

The cyclic behavioral search based pattern mining is defines a time that indicates the stopping criteria for repeatedly performing bidirectional expansion search to efficiently mine frequent web pages in weblog database. The cyclic behavioral pattern mining set the time period t after which reiteration of bidirectional expansion search shall be started periodically after current bidirectional expansion search has been completed. For instance, $t = x$ ms represents that the reiteration of bidirectional expansion search is repeated with period 5 ms after current bidirectional expansion search has been ended. This cyclic behavioral search based pattern mining ends when the time t reaches 5 ms. With help of cyclic behavioral search, BES-CBPM method effectually mines all frequent web pages in web log database which resulting in enhanced mining efficiency. The algorithmic process of cyclic behavioral search based pattern mining is shown in below,

// Cyclic Behavioral Search Based Pattern Mining Algorithm
Input: Amazon Commerce reviews set dataset
Output: Improved Mining Efficiency
Step 1: Begin
Step 2: Cyclic behavioral search assign time period $t = x$
Step 3: **While** ($t = x$) **do**
Step 4: Reiteration of bidirectional expansion search is carried out
Step 5: Efficiently extracts frequent web pages visited by web users
Step 6: **End while**
Step 7:End

Algorithm 2 Cyclic Behavioral Search Based Pattern Mining

As demonstrated in Algorithm 2, at first the cyclic behavioral search based pattern mining algorithm takes Amazon commerce reviews set dataset as input. Then, cyclic behavioral search based pattern mining algorithm defines time period in order to reiterate the bidirectional expansion search for effectually extracting the similar user's interest web pages. The cyclic behavioral search continued until the time t reaches x. Thus, BES-CBPM method enhances the

mining efficiency of web sequential patterns in a significant manner.

3. Experimental Settings

In order to evaluate the effectiveness of proposed, Bidirectional Expansion Search based Cyclic Behavioral Pattern Mining (BES-CBPM) method is implemented in Java language with help of Amazon Commerce reviews set dataset obtained from UCI repository. The Amazon Commerce reviews set dataset [21] comprises of many customer's reviews in Amazon Commerce Website. The proposed BES-CBPM method considers 50 active users (represented by a unique ID and username) who regularly posted comments in these newsgroups. The number of reviews collected for each author is 30 for performing experimental process. The performance of BES-CBPM method is compared against with existing automatic annotation approach [1] and Ant-based clustering algorithm [2]. The efficiency of BES-CBPM method is measured in terms of time complexity, space complexity, mining efficiency, true positive rate of mining and scalability. The experimental process is conducted for many instances with respect to different number of web patterns and averagely ten results is shown in table and graph in order to analyze the proposed BES-CBPM method performance

4. Result and Discussion

In this section, the result of BES-CBPM method is analyzed. The performance of BES-CBPM method is compared against with existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively. The effectiveness of BES-CBPM method is evaluated along with the following metrics with the assist of tables and graphs.

4.1 Measure of Time Complexity

In BES-CBPM method, Time Complexity (TC) measures the amount of time taken for extracting the web sequential patterns from web log database. The execution time is measured in terms of milliseconds (ms) and mathematically represented as,

$$TC = n * time(extracting\ one\ web\ sequential\ pattern) \quad (4)$$

From equation (4), time taken for extracting the web sequential patterns is determined with respect to diverse number of patterns (n). While a Time Complexity is lower, the method is said to be more efficient.

Table 1: Tabulation of Time Complexity

Number of patterns (n)	Time Complexity (ms)		
	Automatic Annotation Approach	Ant-Based Clustering Algorithm	BES-CBPM method
25	0.14	0.12	0.05
50	0.20	0.16	0.09
75	0.28	0.18	0.11
100	0.31	0.21	0.13
125	0.38	0.25	0.18
150	0.45	0.28	0.21
175	0.51	0.32	0.27
200	0.55	0.36	0.32
225	0.56	0.39	0.35
250	0.63	0.45	0.38

Table 1 portrays the tabulation results of time complexity for mining web sequential patterns with respect to diverse number of patterns in the range of 25-250. The BES-CBPM method considers the method with different numbers of web patterns for carried outing the experimental processes using Java language. While considering the 75 web patterns, proposed BES-CBPM method obtains 0.11 ms time complexity to efficiently mine similar user interest web pages whereas automatic annotation approach [1] and Ant-based clustering algorithm [2] obtains 0.28 ms and 0.18 ms respectively. Hence, the time complexity of web sequential pattern mining using proposed BES-CBPM method is lower as compared to other existing works.

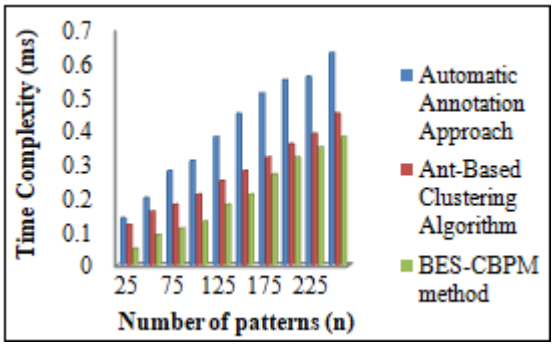

Figure 4: Measurement of Time Complexity versus Number of Patterns

Figure 4 present the impact of time complexity for mining similar user's interest web pages from weblog database versus varied number of patterns in the range of 25-250. As demonstrated in figure, the proposed BES-CBPM method gives better time complexity for extracting frequent web pages browsed by web users from weblog database when compared to automatic annotation approach [1] and Ant-based clustering algorithm [2]. In addition, while increasing the number of web patterns, the time taken for extracting the web sequential patterns is also gets increased for all three methods. But, comparatively the time taken for extracting the web sequential patterns using proposed BES-CBPM method is lower. This is because of usage of bidirectional expansion search algorithm in BES-CBPM method for mining web sequential patterns. The bidirectional expansion search algorithm carried outs two simultaneous search (i.e. forward and backward search) in order to extracts the frequent web pages in weblog. This assists for BES-CBPM method to lessen the time taken for extracting the web sequential patterns from web log database in an effectual manner. Therefore, proposed BES-CBPM method minimizes the time complexity for mining web sequential patterns by 51 % and 29 % when compared to existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively.

4.2 Measure of Space Complexity

In BES-CBPM method, Space Complexity (SC) determines the amount of memory required for storing the mined web

sequential patterns. The space complexity is measured in terms of kilo bytes (KB) and formulated as,

$$SC = N^* \quad memory(storing\ one\ web\ sequential\ pattern) \qquad (5)$$

From equation (5), memory needed for storing the mined web sequential patterns is estimated with respect to different number of patterns. While a space complexity is lower, the method is said to be more effective.

Table 2: Tabulation of Space Complexity

Number of patterns (n)	Space Complexity (KB)		
	Automatic Annotation Approach	Ant-Based Clustering Algorithm	BES-CBPM method
25	260	238	115
50	274	244	122
75	288	249	128
100	299	253	135
125	320	260	141
150	325	268	147
175	336	274	155
200	345	281	162
225	351	290	170
250	360	296	178

Table 2 shows tabulation results of memory space for storing the mined web sequential patterns with respect to dissimilar number of patterns in the range of 25-250 using three methods. While considering the 125 web patterns for accomplishing experimental process, proposed BES-CBPM method attains 141 KB space complexity whereas automatic annotation approach [1] and Ant-based clustering algorithm [2] attains 320 KB and 260 KB respectively. As a result, the space complexity of web sequential pattern mining using proposed BES-CBPM method is lower as compared to other existing works.

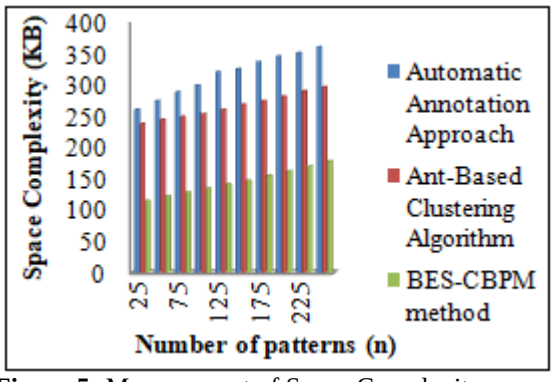

Figure 5: Measurement of Space Complexity versus Number of Patterns

Figure 5 depicts the impact of space complexity for storing the extracted web sequential patterns versus various numbers of patterns in the range of 25-250. As illustrated in figure, the proposed BES-CBPM method gives minimum space complexity for mining the similar user's interest web pages from web log database when compared to automatic annotation approach [1] and Ant-based clustering algorithm [2]. Further, while increasing the number of web patterns, memory space taken for storing the extracted web sequential patterns is also gets increased for all three methods. But, comparatively the memory space taken for storing the extracted web sequential patterns using proposed BES-CBPM method is lower. This is owing to usage of bidirectional expansion search algorithm in BES-CBPM method. The bidirectional expansion search algorithm performs two simultaneous searches with objective of extracting only the similar users interest web pages in weblog database. This supports for BES-CBPM method to lessen the memory space consumed for storing the web sequential patterns in a significant manner. Thus, proposed BES-CBPM method reduces the space complexity of web sequential patterns mining by 55 % and 46 % when compared to existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively.

4.3 Measure of Mining Efficiency

In BES-CBPM method, the Mining Efficiency (ME) measures the ratio of the number of web patterns mined as similar user's interest to the total number of web patterns taken as input. The Mining Efficiency is determined in terms of percentages (%) and formulated as,

$$ME = \frac{number\ \ of\ web\ patterns\ \ mined\ \ as\ similar\ \ user's\ interest}{total\ \ number\ \ of\ web\ \ patterns} * 100 \qquad (6)$$

From equation (6), mining efficiency of web sequential pattern is determined with respect to different number of web patterns. While a mining efficiency of web sequential patterns is higher, the method is said to be more effectual.

Table 3: Tabulation of Mining Efficiency

Number of patterns (n)	Mining Efficiency (%)		
	Automatic Annotation Approach	Ant-Based Clustering Algorithm	BES-CBPM method
25	68.23	71.26	88.23
50	70.12	73.56	87.56
75	71.65	75.16	88.23
100	73.52	76.98	89.12
125	74.69	78.92	89.95
150	76.98	81.02	90.65
175	78.23	82.13	91.03
200	79.26	83.26	91.89
225	80.13	84.03	93.21
250	81.08	84.83	95.16

The tabulation results of mining efficiency for web sequential pattern with respect to various numbers of patterns in the range of 25-250 using three methods is demonstrated in Table 3. While considering the 150 web patterns for conducting experimental work, proposed BES-CBPM method achieves 90.65 % mining efficiency whereas automatic annotation approach [1] and Ant-based clustering algorithm [2] achieves 76.98 % and 81.02 % respectively. Therefore, the mining efficiency of web sequential pattern mining using proposed BES-CBPM method is higher as compared to other existing works.

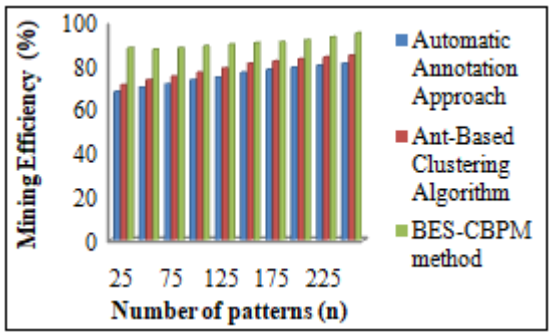

Figure 6: Measurement of Mining Efficiency versus Number of Patterns

Figure 6 describes the impact of mining efficiency versus diverse number of patterns in the range of 25-250. As depicted in figure, the proposed BES-CBPM method gives higher mining efficiency for extracting web sequential patterns from web log database when compared to automatic annotation approach [1] and Ant-based clustering algorithm [2]. Besides, while increasing the number of web pattern, mining efficiency of similar users interest is also gets increased for all three methods. But, comparatively the mining efficiency using proposed BES-CBPM method is higher. This is due to usage of bidirectional expansion search algorithm and cyclic behavioral search based pattern mining in BES-CBPM method. The bidirectional expansion search algorithm mines the frequent web pages in weblog database. Some of frequent web pages in weblog database may not be mined during the Bidirectional Expansion Search. Hence, BES-CBPM method applied cyclic behavioral search based pattern mining for effective web sequential mining. This helps for BES-CBPM method to improve the mining efficiency of web sequential patterns in an efficient manner. Hence, proposed BES-CBPM method increase the mining efficiency by 20 % and 15 % when compared to existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively.

4.4 Measure of True Positive Rate of Mining

In BES-CBPM method, True Positive Rate (TPR) of mining is defined as the ratio of number of correctly mined web access patterns as similar users interest to the total number of web access patterns taken as input. The true positive rate of mining is measured in terms of percentages (%) and represented as,

$$TPR = \frac{correctly\ \ mined\ \ web\ sequential\ \ patterns\ \ from\ web\ log\ as\ similar\ \ users\ interest}{total\ \ number\ \ of\ web\ patterns} * 100 \quad (7)$$

From equation (7), number of correctly mined web access patterns as similar user's interest is determined with respect to diverse number of patterns. When a true positive rate of web sequential pattern mining is higher, the method is said to be more effective.

Table 4: Tabulation of True Positive Rate of Mining

Number of patterns (n)	True Positive Rate of Mining (%)		
	Automatic Annotation Approach	Ant-Based Clustering Algorithm	BES-CBPM method
25	56.13	70.35	81.26
50	57.16	72.35	83.26
75	59.32	74.91	84.16
100	62.05	76.31	86.98
125	63.97	78.23	89.13
150	66.12	80.03	89.95
175	69.01	81.95	91.32
200	70.15	83.62	92.65
225	72.06	84.06	94.26
250	75.69	87.65	96.37

The tabulation result of true positive rate for mining web sequential patterns with respect to diverse numbers of patterns in the range of 25-250 using three methods is illustrated in Table 4. While considering the 175 web patterns, proposed BES-CBPM method gets 91.32 % true positive rate whereas automatic annotation approach [1] and Ant-based clustering algorithm [2] acquires 69.01 % and 81.95 % respectively. For that reason, the true positive rate using proposed BES-CBPM method is higher as compared to other existing works.

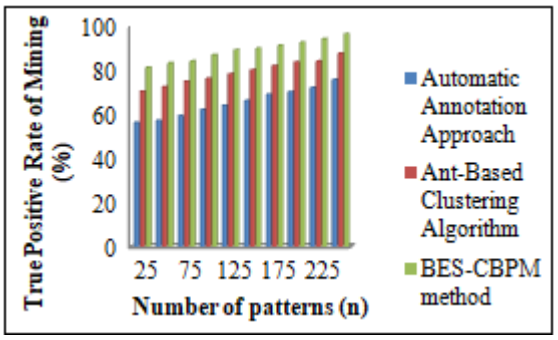

Figure 7: Measurement of True Positive Rate of Mining versus Number of Patterns

Figure 7 demonstrates the impact of true positive rate for mining similar user's interest web pages versus various numbers of patterns in the range of 25-250. As exposed in figure, the proposed BES-CBPM method gives higher true positive rate for extracting frequent web pages visited by web users in web log database when compared to automatic annotation approach [1] and Ant-based clustering algorithm [2]. Moreover while increasing the number of web patterns, true positive rate for mining similar user's interest web pages is also gets increased for all three methods. But, comparatively the true positive rate using proposed BES-CBPM method is higher.

This is due to application of bidirectional expansion search algorithm in BES-CBPM method in which similarity measurement is used to significantly discover and mine similar user's interest web pages. This assists for BES-CBPM method to precisely extract web access patterns as similar user's interest. Hence, proposed BES-CBPM method increase the true positive rate of mining web sequential patterns by 20 % and 13 % when compared to existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively.

4.5 Measure of Scalability

Scalability measure the capability of BES-CBPM method to handle huge size of weblog database for mining the web sequential patterns effectively. While the scalability of web sequential mining is higher, the method is said to be more effectual.

Table 5: Tabulation for Scalability

Methods	Scalability (%)
Automatic Annotation Approach	69.15
Ant-Based Clustering Algorithm	79.31
BES-CBPM method	89.25

The tabulation result of scalability for mining frequent web pages visited by web users' using three methods is portrayed in Table 5. From the table value, it is expressive that the scalability rate of frequent web pages visited by web users' using proposed BES-CBPM method is higher as compared to other existing automatic annotation approach [1] and Ant-based clustering algorithm [2].

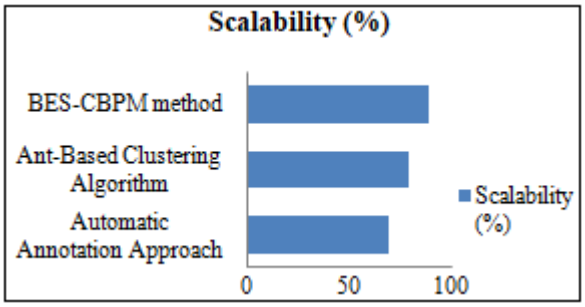

Figure 8: Measurement of Scalability versus Number of Patterns

Figure 8 illustrates the impact of scalability for mining similar user's interest web pages while increasing the size of input web patterns using three methods. As revealed in figure, the proposed BES-CBPM method gives higher scalability rate for efficiently mining the frequent web pages browsed by web users when compared to automatic annotation approach [1] and Ant-based clustering algorithm [2]. Also, while increasing the number of web patterns, scalability rate for extracting similar user's interest web pages is also gets increased for all three methods. But comparatively the scalability using proposed BES-CBPM method is higher. This is owing to application of bidirectional expansion search algorithm in BES-CBPM method. The bidirectional expansion search algorithm is mainly designed for exponentially growing size of weblog database. With application of bidirectional expansion search algorithm, bidirectional expansion search algorithm effectively mines the frequent sequential patterns while increasing the input size of web patterns. This assists for BES-CBPM method to improve the scalability rate in an effective manner. Therefore, proposed BES-CBPM method increase the scalability of mining frequent web pages by 23 % and 13 % when compared to existing automatic annotation approach [1] and Ant-based clustering algorithm [2] respectively.

5. Related Works

Apriori prefix tree (PT) algorithm was intended in [11] for predicting and mining the patterns of user's visit web pages. However, the efficiency of this algorithm was not effectual. High Utility-Probability Sequential Pattern Mining (HUPSPM) was intended in [12] for extracting sequential patterns in uncertain sequence databases and faster the mining process. However, time complexity of HUPSPM was not at required level. A new approach based on a user-defined scoring mechanism was introduced in [13] to mine patterns from web log database through combining clustering with a heuristic-based pattern extraction algorithm. But, accuracy of mining web sequential patterns was poor. A MapReduce-based algorithm was presented in [14] for find outing navigation patterns of web users through enhancing performance of frequent-sequence-pattern-mining algorithms. However, time and space complexity was more.

Time based discovering of web user patterns was introduced in [15] to optimally extracting the logged web usage data of users. However, scalability rate was remained unsolved. A novel approach was designed in [16] for effectively mining web log and for online navigational behavior prediction. But, mining accuracy was not enough. Pattern-Based Web Mining Using Data Mining Techniques was presented in [17] to mine information from Web data. However, the performance of pattern-based web mining not effectual which lacks mining efficiency of web patterns. An efficient Apriori algorithm was developed in [18] with objective of mining user frequent patterns from weblog database. But, mining efficiency of frequent patterns was poor. The difficulty of mining probabilistically frequent sequential patterns (p-FSPs) in uncertain databases was solved in [19]. However, mining frequent sequential patterns in weblog database was not considered. A Web Log Mining method was presented in [20] for extracting useful patterns from the web log data and web recommendation and personalization. However, mining efficiency was not efficient.

6. Conclusion

An efficient Bidirectional Expansion Search based Cyclic Behavioral Pattern mining (BES-CBPM) method with goal of improving the mining efficiency of frequent web pages in weblog with minimum time and space complexity. The objective of BES-CBPM method is achieved with application of Bidirectional Expansion Search algorithm and cyclic behavioral search based pattern mining. The BES-CBPM method initially employed Bidirectional Expansion Search algorithm for extracting the maximum visited web pages by web users from weblog database. The Expansion Search algorithm formulated in BES-CBPM method performs two searches at the same time with objective of extracting the similar user's interest web pages. This in turn significantly minimizes time and space complexity for mining the web sequential patterns. Additionally, BES-CBPM method utilized cyclic behavioral search based pattern mining in order to achieve very mining efficiency for extracting frequent web pages. The cyclic behavioral search based pattern mining used in BES-CBPM method repeats the bidirectional expansion search for a specific period of time to increase mining performance. This cyclic behavioral search process helps for BES-CBPM method to extract the missed frequent web pages during bidirectional expansion search. Therefore, proposed BES-CBPM method attains very high mining efficiency. The performance of BES-CBPM method is test with the parameters such as time complexity, space complexity, mining efficiency, true positive rate and scalability. With the experiments performed for BES-CBPM method, it is expressive that the mining efficiency presents more precise results for mining the web sequential pattern for weblog database as compared

to state-of-the-art works. The experimental results shows that BES-CBPM method is web sequential pattern provides better performance with an enhancement of mining efficiency and the reduction of time complexity when compared to the state-of-the-art works.

References

[1] Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng, Clement Yu, "Annotating Search Results from Web Databases", IEEE Transactions On Knowledge And Data Engineering, Volume 25, Issue 3, Pages 514 – 527, March 2013.
[2] D. Anandhi, M. S. Irfan Ahmed, "Prediction of user's type and navigation pattern using clustering and classification algorithms", Cluster Computing, Springer, Pages 1–10, 2017.
[3] Xiuming Yu, Meijing Li, Taewook Kim, Seon-phil Jeong, and Keun Ho Ryu1, "An Application of Improved Gap-BIDE Algorithm for Discovering Access Patterns", Hindawi Publishing Corporation, Applied Computational Intelligence and Soft Computing, Volume 2012, Article ID 593147, Pages 1-7, 2012.
[4] Kuo-Wei Hsu, "Efficiently and Effectively Mining Time-Constrained Sequential Patterns of Smartphone Application Usage", Hindawi, Mobile Information Systems, Volume 2017, Article ID 3689309, Pages 1-18, 2017.
[5] Binu Thomas and G. Raju, "A Novel Web Classification Algorithm Using Fuzzy Weighted Association Rules", Hindawi Publishing Corporation, ISRN Artificial Intelligence, Volume 2013, Article ID 316913, Pages 1-10, 2013.
[6] Sheetal Sharma, Swati singh lodhi, "Development of Decision Tree Algorithm for Mining Web Data Stream", International Journal of Computer Applications, Volume 138, Issue 2, Pages 34-43, March 2016.
[7] NazliMohd Khairudin, AidaMustapha, andMohd Hanif Ahmad, "Effect of Temporal Relationships in Associative Rule Mining for Web Log Data", Hindawi Publishing Corporation, The Scientific World Journal, Volume 2014, Article ID 813983, Pages 1-10, 2014.
[8] Priyanka Baraskar, Supriya Chavan, Dipshree Dhage, Samruddhi Giri, Jayashree Jha, "Extracting Frequent Sequences from Web Log Data using Sequence Tree Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Volume 4, Issue 3, Pages 10-12, March 2015.
[9] Gopalakrishna Kurup Raju, Achuthan Nair Rajimol, "A Novel Weighted Support Method for Access Pattern Mining", International Arab Journal of e-Technology, Volume 3, Issue 4, Pages 201-209, June 2014.
[10] Osvaldo Navarro, Rene Cumplido, Luis Villasen or-Pineda, Claudia Feregrino-Uribe, Jesu s Ariel Carrasco-Ochoa, "A Node Linkage Approach for Sequential Pattern Mining", PLOS One, Volume 9, Issue 6, Pages 1-16, 2014 .
[11] R Geetharamani, P Revathy And Shomona G Jacob, "Prediction of users webpage access behaviour using association rule mining", Indian Academy of Sciences, Springer, Volume 40, Issue 8, Pages 2353–2365, December 2015.
[12] Binbin Zhang, Jerry Chun-Wei Lin, Philippe Fournier-Viger, Ting Li, "Mining of high utility-probability sequential patterns from uncertain databases", PLoS ONE, Volume 12, Issue 7, Pages 1-21, 2017.
[13] Oznur Kirmemis Alkan, Pinar Karagoz, "WaPUPS:Web access pattern extraction under user-defined pattern scoring", Journal of Information Science, Volume 42, Issue 2, Pages 1-13, 2016.
[14] Meijing Li, Xiuming Yu, Keun Ho Ryu, "MapReduce-based web mining for prediction of web-user navigation", Journal of Information Science, Volume 40, Issue 5, Pages 557-567, 2014.
[15] Emrah Donmez, Alper Ozcan, "Time Based Discovering Of Web User Patterns to Optimize Web Sites and Hyperlinks", International Journal of Advanced Computational Engineering and Networking, Volume 3, Issue 8, Pages 14-20, 2015.
[16] Abdelghani Guerbas, Omar Addam, Omar Zaarour, Mohamad Nagi, Ahmad Elhajj, Mick Ridley, Reda Alhajj, "Effective web log mining and online navigational pattern prediction", Knowledge-
[17] Sheng-Tang Wu and Yuefeng Li, "Pattern-Based Web Mining Using Data Mining Techniques", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol ume 3, Issue 2, Pages 163-167, April 2013.
[18] S.VijayaKumar, A.S.Kumaresan, U.Jayalakshmi, "Frequent Pattern Mining in Web Log Data using Apriori Algorithm", International Journal of Emerging Engineering Research and Technology, Volume 3, Issue 10, Pages 50-55, October 2015.
[19] Zhou Zhao, Da Yan and Wilfred Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 5, Pages 1171 – 1184, May 2014.
[20] Jaymin Desai, Risha Tiwari, "Web Log Mining Using Multiitem Sequntial Pattern Based On PLWAP", International Journal For Technological Research In Engineering, Volume 4, Issue 7, Pages 1015-1016, March-2017.
[21] Amazon Commerce reviews set Data Set: https://archive.ics.uci.edu/ml/datasets/Amazon+Commerce+reviews+set

Author Profile

A.P.Selva Prabhu received the B.Sc and M.Sc degree in Computer Science from Bharathiar University, Tamilnadu, India, in 2005 and 2007, and the M.Phil degree in Computer Science from Bharathiar University in 2011. He is a Assistant Professor in Computer Science at Bharathiar University Arts and Science College, Modakkuruchi, Erode, Tamilnadu, India.

T. Ravichandran received the B.E Degree in Electronics and Communication Engineering from Bharathiar University, Tamilnadu, India, in 1994 and the M.E. degree in Computer Science and Engineering from Madurai Kamaraj University, Tamilnadu, India, in 1997 and the Ph.D. degree from Periyar University, Tamilnadu, India, in 2007. He is a professor and dean in the department of Electronics and Communication Engineering at SNS College of Technology,

Coimbatore, Tamilnadu, India. Under his supervision 30 research scholars completed their Ph.D and he has received many awards including best faculty award from KSR College of technology. He is a reviewer of many international journals and also an editorial member of many reputed journals.