International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Web Personalization using Amalgamation of Web Navigational Patterns & User Profiles

Sowbhagya M. P., Yogish H. K., G. T. Raju

Submitted: 26/08/2023 **Revised**: 14/10/2023 **Accepted**: 27/10/2023

Abstract: Web personalization proposes customized Web sites to users or anticipates personalised Web things to them based on their individual discoveries regarding user profiles or navigational patterns. User profiles are fundamentally made to display particular user guiding habits that were found through web usage research. The frequent sequential patterns that are derived from online usage data utilizing frequent sequential pattern mining techniques are what are known as navigational patterns. In this research, a novel method for online personalization is developed using the information from user profiles and navigational habits. A set of active navigational patterns, user profiles, and the prediction period are first read as input. Next, user profiles and navigational patterns are compared to determine the anticipated pages. Then, considering the most crucial user characteristics and navigational patterns, each page's ranking is calculated. The best top n-pages are then recommended. Two data sets from the KDDCUP and scholarly websites were used in the testing. The findings demonstrate that the suggested method successfully provides web users with a wealth of data for anticipating and recommending tailored web pages. The suggested approach permits a 6.3-fold increase in traffic that is tolerable with a maximum latency saving ratio of 7.5.

Keywords: *Web Personalization, Navigational pattern, User Profiles*

1. Introduction

It's fantastic that the Internet has developed into a reliable platform for information storage, dissemination and recovery. Web users consistently experience the negative effects of information overload and suffocation concerns. Customers should therefore pay close attention to issues like modest precision and review rate when looking for important information online. However, modern data mining methods can be applied to the enormous amount of data and information that is present on the Internet to unearth a sizable number of extremely pertinent pieces of information.

First, information is viewed in the form of a set of dynamic navigational patterns, set of client profiles, and an anticipated time frame. The pages are then predicted by examining how similar client profiles and navigational patterns are. The most crucial customer traits and navigational patterns are then chosen to establish the rating for each page. Finally, internet visitors are advised to visit the top n-pages with the highest rank.

The relevant work is revealed in Section 2. Section 3 presents a suggested system architecture. A novel approach to online personalization that includes information from user profiles and browsing patterns is presented in section 4. Section 5 provides experimental data and analyses. With room for more research, Section 6 offers the paper's conclusion.

[1] *Dayananda Sager Academy of Technology and Management,Bangalore*
Sowbhagya.mp@gmail.com
[2] *M S Ramaiah Institute of Technology,Bangalore*
Yogishhk@gmail.com
[3] *S J C Institute of Technology,Chikkaballapur*
gtraju1990@yahoo.com

2. Related Work

If you In the relevant literature, the problem of making suggestions to website users has attracted a lot of attention. The majority of web personalization research efforts are in line with the development of in-depth web usage mining research, which solely considers the navigational habits of (anonymous or registered) website visitors [1, 2 and 3]. However, pure usage-based customisation has several drawbacks. For example, when there is insufficient usage data to identify patterns relating to specific navigational actions, or when the website's content is updated and new pages are introduced but are not yet recorded in the web logs. Such systems are also extremely vulnerable to the training data that was used to build the predictive model, given the temporal features of the web's usage. As a result, some study approaches employ data from additional sources, such as web content [4, 5 and 6] or web structure [7, 8], in order to improve the web personalization process. We must remember that the internet is more than just a database of documents that people peruse. The internet is a directed labeled graph with a vast number of links connecting its web pages. The structural properties of the online graph as well as the underlying semantics of the web pages and hyperlinks play a significant and determining role in how users navigate. Numerous studies have developed frameworks that translate users' navigational behavior into ontologies and incorporate this information into collaborative filtering systems [10], Markov model-based recommendation systems [4], or semantic web sites [9]. With the express purpose of improving framework tuning and method assessment for client considerations, the authors of [11] demonstrated and examined the usefulness of brief, understandable client profiles. A framework based on the idea of using personalization of web information is offered in [12], which includes an examination of a few customised methods for gathering web data and correlations between them. The two features were utilised by the [13] creators to meet a client's transient necessity. The first step is to gradually increase each client's degree of enthusiasm as time passes. Second, an action was noted in online browsing logs. Paper[15] has presented a recommendation method that suggests a list of URLs primarily based on past behavior of the client. User profiles and navigational behaviours must therefore be incorporated for efficient online personalisation. This study's primary goal is to suggest a novel method for combining user profiles and navigational patterns in order to increase the overall efficiency of the web personalization process.

3. Proposed System Architecture

This study recommends using client-side proxies to combine user profiles & navigational patterns for online personalization and caching. Figure 1 depicts the suggested system design. To prevent frequent round-trip delays between Web clients and the origin Web servers, the proxy is set up close to the clients. A prediction engine is used by the origin Web server in our architecture to forecast the top-N pages that should be suggested for customisation. In Figure 1, the proxy is shown responding to queries from Web clients. The proxy will send the request back to the originating Web server for processing if it doesn't fit in the cache. The requested page will be fetched from the Web page repository once the origin server logs the request into a record and scans the top-N pages. If a few of the top-N pages respond to this request, they will be piggybacked onto the response message as recommendations and returned to the proxy. Once the proxy receives the answer from the origin Web server, which contains the piggybacked hints, it will decide whether or not to cache the piggybacked inferred sites before giving the client the page they requested. We assume that the proxy will maintain contact with the origin Web server to obtain the personalization hints connected with that request after the proxy has sent the client the response if a cache hit is discovered (the client's request can be fulfilled directly by the proxy's local cache). As a result, we may check the customization recommendations from the origin Web server to make sure they are always current for each request.

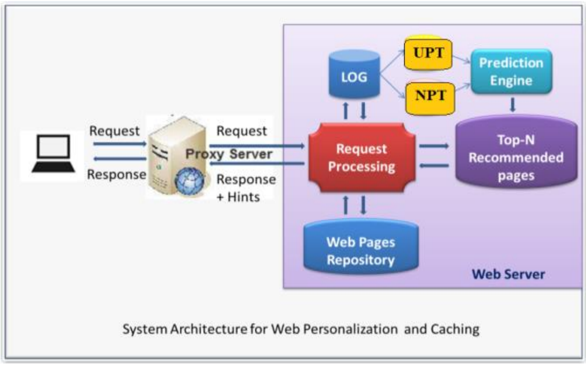

Fig 1: Web personalization & caching system architecture

4. A New Algorithm for Personalizing Websites

Usually, a web recommendation's objective is to foresee and personalize web presentations in a way that visitors will find appealing based on their preferences. There are two methods for achieving this goal. On the one hand, we may forecast the preferred information for this particular user by taking into account the navigational behaviors of the currently active users. On the other hand, by determining which users' access patterns are most similar to those of the currently active user, we can provide the customised Web content. The other side, we can recommend the customized Web content by identifying the access behaviors of other users that are most comparable to the current active user. We use a model-based approach in this work's Web recommendation system. As mentioned in our earlier study, efficient FSP mining methods are used to extract common sequential patterns from web activity data. These serve as the patterns for navigation. We create user profiles from the identified user session clusters. We present the RPUPBP algorithm, which integrates an individual Profile Table (UPT) and Navigational Pattern Table (NPT) given a UPT and NPT to predict and suggest Web pages for an individual or group. We use the well-known cosine function to assess how much the user profile and navigational pattern match one another. Next, we choose the most crucial user profile and navigational pattern. Before concluding, we produce the top-N suggestion sites. The following is a full description of the process.

Algorithm [RPUPBP]*:* Web page suggestions for a user or group are based on the integration of the User Profile Table (UPT) and the Navigational Pattern Table (NPT).

Input: Currently used navigational patterns NPT as well as a no. of user profiles UPT and the anticipated timeframe τ = {Weekly, Fortnightly, and Monthly}

Output: Set of top-N pages predicted / recommended for the user/group within τ

Steps:

1. If new user? Then extract the *BPs* and build the *UP*. Identify the group for which user belongs to and update the corresponding UPT and NPT.
2. For each user/group, read the *UPT* and *NPT* for the prediction period τ

The *BPs* and *UPs* are modeled as n-dimensional vectors over the page space within the web site. For example, k^{th} BP and UP are given as follows:

$BP_k=[a_1,a_2,a_3,\ldots,a_n]$ a_i = 1- if accessed already, 0 – otherwise for pages $p_1,p_2,p_3,\ldots,p_n$

$UP_k=[w_1,w_2,w_3,\ldots,w_n]$ w_i represent the weight/frequency of the page p_i in UP_k

3. Predict the pages by measuring similarity between BPs and Ups

$$Similarity\ (BP_k, UP_k) = (BP_k . UP_k)/(\ \|BP_k\|_2\ \|UP_k\|_2)$$

$$where\ \ BP_k . UP_k = \sum_{i=1}^{k} w_i^k . a_i^k,$$

$$\|BP_k\|_2 = \sqrt{\sum_{i=1}^{k} (a_i^k)^2}$$

and $\|UP_k\|_2 = \sqrt{\sum_{i=1}^{k} (w_i^k)^2}$

4. Choose/Select the BP and UP with maximum similarity called the most significant BP and UP (BP_s, UPs)

(BP_s, UPs) = Max($Similarity\ (BP_k, UP_k)$) for k = 1,2,..,n

5. From the selected BP_s *and* UPs, calculate the *rank* $r(p_i)$ for each page p_i

$r(p_i) = \sqrt{Similarity\ (BP_s, UP_s) * w_i^s}$

Thus each page is assigned a *rank* between 0 and 1. Note that the *rank* will be 0 if the page is already visited in the current session

6. *By sorting the rankings in descending order, selecting the top-N pages with the highest ranks, and so forth, create the top-N propose pages RPs.*

$RPs = \{\ p_i^s \mid r(p_i^s) > r(p_{i+1}^s), for\ i = 1,2,3,\ldots,n-1\ \}$

5. Results

In order to assess the efficacy of the recommended approach, experiments have been carried out on two real-world data sets collected from the log files of the KDDCUP (www.ecn.purdue.edu/kddcup/) and ACADEMIC (rnsit.ac.in) websites. When preprocessing is complete, KDDCUP will have 69 pages and 6305 user sessions. These data sets are shown as usage matrices, with each column denoting a page and each row denoting a session. These matrices provide as a source of initial data from which usage data and dormant semantic links are recovered. As shown in Tables 1 and 2, each user profile is made up of a few key pages, and the pertinent assistance is expressed in a normalised form. Table 1 shows two user profiles produced using the KDD dataset from our prior study.

Table 1: User profile examples made using the KDDCUP dataset

Profile#	Page #	Page title	Support
Profile -1	29	/Main-shopping_cart.html	1.00
	4	/Products-productDetailleagwear.html	0.86
	27	/Main-Login2.html	0.67
	8	/Main-home.html	0.53
	44	/Check-express_Checkout.html	0.38
	65	/Main-welcome.html	0.33
	32	/Main-registration.html	0.32
	45	/Checkout-confirm_order.html	0.26
Profile -2	11	/Main-vendor2.html	1.00
	8	/Main-home.html	0.40
	12	/Articles-dpt_about.html	0.34
	13	/Articles-dpt_about_mgmtteam.html	0.15
	14	/Articles-dpt_about_broadofdirectors.html	0.11

Table 2: An example of a user profile made with the help of information from the Academic website

Profile#	Page #	Page title	Support
Profile - 1	3	/Admissions.html	1.00
	6	/Placement.html	0.41
	9	/sports&culture.html	0.24
	28	/awards.html	0.21
	29	/accolades.html	0.11
	46	/aboutus.html	0.11
Profile - 2	14	/campus.html	1.00
	7	/hostel.html	0.35
	5	/library.html	0.32
	30	/bestteachers.html	0.13
	1	/index.html	0.11
Profile - 3	1	/index.html	1.00
	12	/courses.html	0.78
	21	/pp_cse.html	0.40
	35	comp-visited.html	0.17
	6	/Placement.html	0.12

The same three profiles are obtained from Table 2. The produced profiles reveal that the majority of users navigate in a specific way, while just a small number of users have multiple interests.

Sample navigational patterns taken from the Academic web site dataset are displayed in Table 3.

Table 3. Sample Navigational Patterns(FSPs) extracted from ACADEMIC web sitedataset

No. of User	No. of FSPs – Page Ids	No. of User	No. of FSPs – Page Ids
U_{23}	P_{13}, P_{14}	U_{67}	P_7,P_5
	P_{23}, P_{14}		P_7,P_{22}
	P_{23}, P_{13}		P_5,P_{22}
U_{10}	P_{23},P_5		P_5, P_5
	P_{23}, P_{18}		P_{22}, P_7
	P_5 ,P_{18}		P_{22}, P_5
U_{32}	P_1,P_9	U_7	P_{22} ,P_4
	P_1 ,P_{14}		P_{22},P_{21}
	P_9 ,P_{14}		P_4,P_{21}
	P_9, P_1		P_4, P_{22}
	P_{14}, P_1		P_{21}, P_{22}
	P_{14}, P_9		P_2, P_4
U_{68}	P_7,P_5	U_{55}	P_{42},P_8
	P_7,P_{23}		P_{42},P_4
	P_5 ,P_{23}		P_8 ,P_4
	P_5, P_7		P_8, P_{42}
	P_{23}, P_7		P_4, P_{42}
	P_{23}, P_5		P_4, P_8

5.1. Metrics for Performance Evaluation:

Utilizing key performance indicators for personalization costs, prediction accuracy, and perceived latencies among users, the algorithm's performance has been assessed. For instance, when a prediction is made by the algorithm and personalization is implemented in the actual system. Because of this, each prediction index has a dual index; for example, we can talk about the precision of both the prediction engine and the personalization process.

- *Recall (Rc): Proportion of user-requested pages that were tailored in advance.* The recall determines the proportion of user-requested pages that were anticipated and customized beforehand.

  Rc = No. of hits/ Numerous user requests

- *Precision (Pc): Percentage of all forecasts (or custom pages) that were correct. The precision measures the ratio of hits, or the number of pages that were expected and adapted, to the overall number of hits.*

  Pc = No. of hits / No. of Predicted

- *Latency to page ratio (Lp:) compares the latency with and without customisation to determine the difference in latency.* The difference in latency between customisation and non-personalization is calculated using the latency per page ratio (Lp). The results of a baseline experiment without personalisation, which examined the page delay, provided as a benchmark for comparison in the investigations. The following is how the page latency (PL) is calculated:

  ∆PL = Avg. Page Latency with Personalization / Avg. Page Latency without personalization

Similar calculations are used to get the page latency saving percentage (PL(%)):

*∆PL(%) = (1 - Average Page Latency with Personalization / Average Page Latency without personalization) * 100*

Because the goal of this work is to investigate the greatest advantage seen by web users, the main presentation metric for assessing the efficacy of prediction and personalisation is page latency saving.

- Increasing traffic (Tr) is the ratio of the amount of data carried over the network

when customization is used to the amount of data transferred when it is not. The additional bytes produced by tailored pages that are never requested by the user make up the increased traffic.

$$\Delta Tr_B = (\text{Pages not used}_B + \text{Network overhead}_B + \text{User requests}_B) / \text{User requests}_B$$

Figures 2, 4, and 6 illustrate the accuracy & recall for the weekly, fortnightly, and monthly time windows, respectively. Similar to this, Figures 3, 5 & 7 correspondingly illustrate thee latency of page(PL) and the page latency saving percentage (PL(%) for time windows of weekly, fortnightly, and monthly.

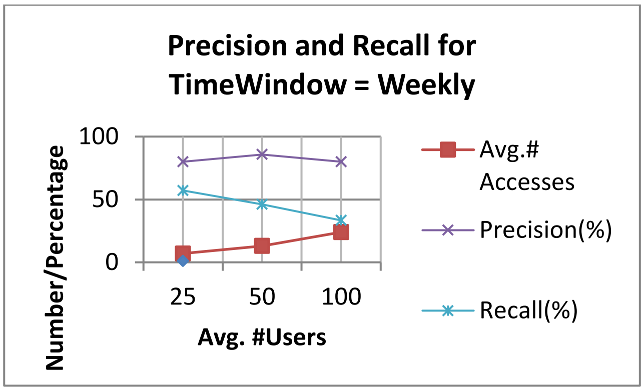

Fig 2: Precision and Recall for the Weekly Time Window

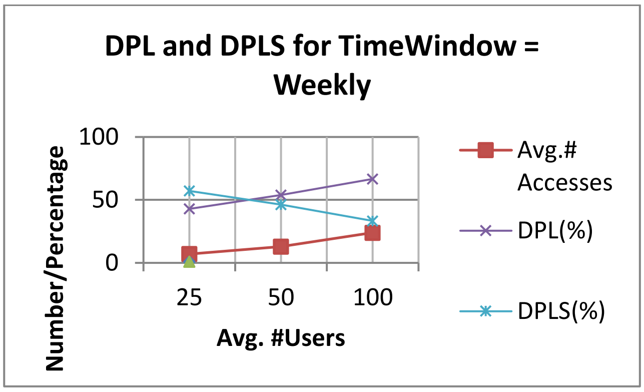

Fig 3: For TimeWindow=Weekly, DPL and DPLS

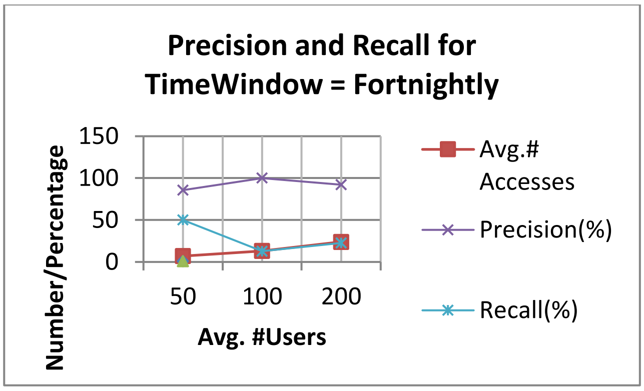

Fig 4: For TimeWindow=Fortnightly, precision and recall

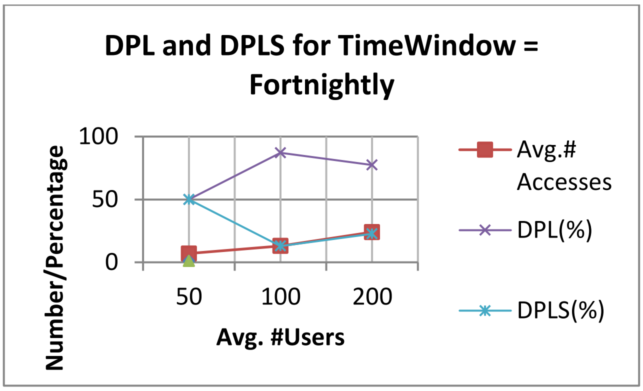

Fig 5: For TimeWindow=Fortnightly, DPL and DPLS

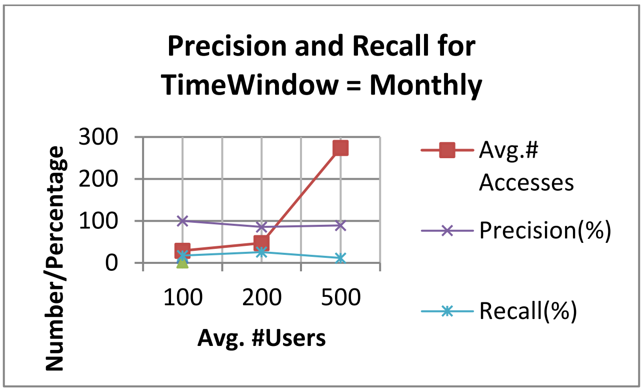

Fig 6: Monthly Precision and Recall Time Window

Table 4 contains the information for the weekly, fortnightly, and monthly window sizes, including the average user count, accesses, projected pages, recall, precision, and latency. Data on network traffic growth and the percentage of latency saved are shown in Table 5.

Table 4. Information about averages for users, accesses, projected pages, recall, precision, and delay Weekly, biweekly, and monthly time windows

Period	Avg. No. Users	Avg. No. Accesses	Web Personalization							
			No. of Pages Predicted	No. of Hits	Recall (%)	Precision (%)	Avg. Lp	Avg. LNp	DPL(%)	DPLS(%)
Weekly	25	7	5	4	57.14	80.00	3	7	42.86	57.14
	50	13	7	6	46.15	85.71	7	13	53.85	46.15
	100	24	10	8	33.33	80.00	16	24	66.67	33.33
Fortnightly	50	12	7	6	50.00	85.71	6	12	50.00	50.00
	100	31	4	4	12.90	100.00	27	31	87.10	12.90
	200	53	13	12	22.64	92.31	41	53	77.36	22.64
Monthly	100	29	5	5	17.24	100.00	24	29	82.76	17.24
	200	47	14	12	25.53	85.71	35	47	74.47	25.53
	500	274	36	32	11.68	88.89	242	274	88.32	11.68

Table 5. Statistics demonstrating a decrease in latency and an increase in network traffic

Storage Size	Avg. Lp	Avg. LNp	LSR		Network Traffic Increase	
			Without Personalization	With Personalization	Without Personalization	With Personalization
1	3	7	42.86	47.14	57.14	52.86
2	6	12	50.00	55.00	50.00	45.00
4	7	13	53.85	59.23	46.15	40.77
8	16	24	66.67	73.33	33.33	26.67
16	41	53	77.36	85.09	22.64	14.91
32	27	31	87.10	95.81	12.90	4.19

For TimeWindow=Monthly, Figure 8 shows the Latency Save Ratio vs. Cache Size. Similar to Figure 8 but with TimeWindow=Monthly, Figure 9 displays Network Traffic Growth v/s Cache Size. The Hit Ratio vs. Cache Size for TimeWindow=Fortnightly and Monthly, respectively, is shown in Figures 10 and 11. Figures 8 and 9 show the greatest latency saving ratio and 6.3 percent traffic growth, respectively, that our technology is able to achieve.

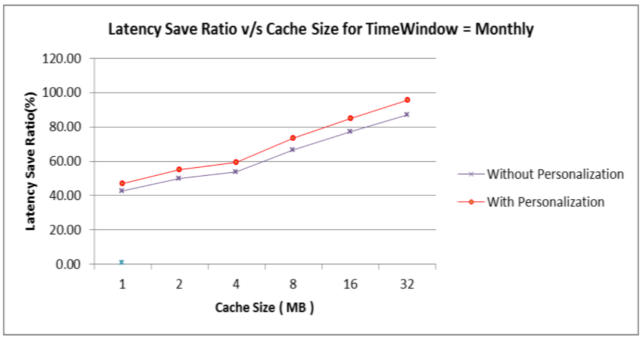

Fig 8: TimeWindow=Monthly Cache Size vs. Latency Save Ratio

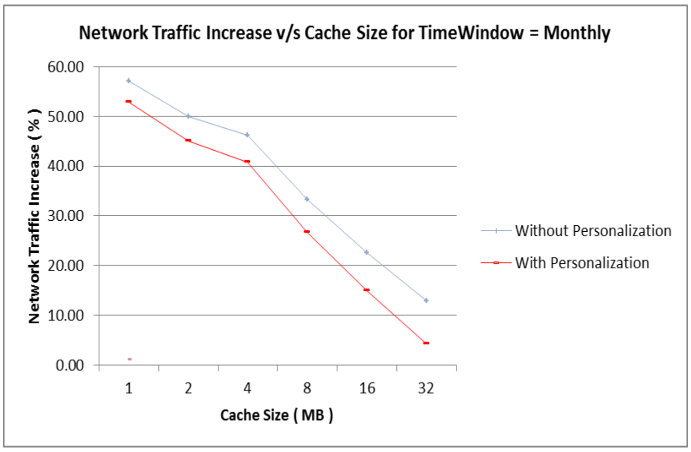

Fig 9: Increase in Network Traffic vs. Cache Size for Time Window = Monthly

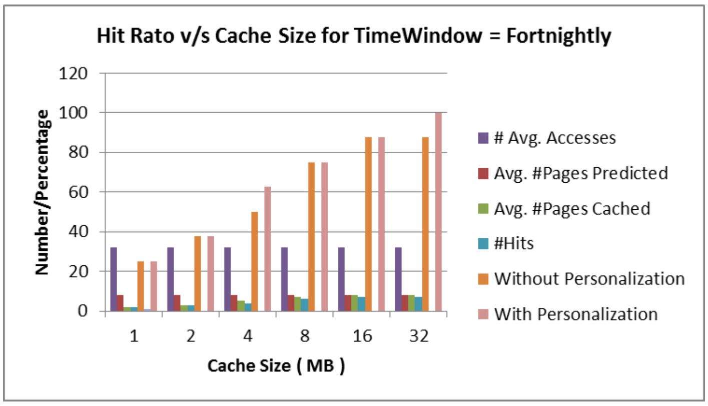

Fig 10: Hit Ratio v/s Cache Size for *TimeWindow*=Fortnightly

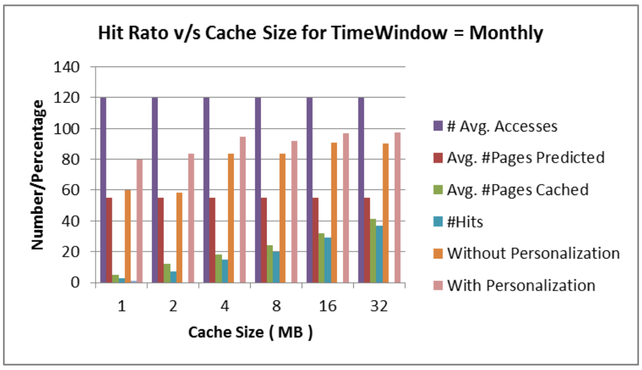

Fig 11: Hit Ratio v/s Cache Size for ***TimeWindow***=Monthly

6. Conclusion

A website's user interface is content-driven. Users typically look for products or services related to a specific topic. Therefore, key factors in the process of online personalization should include the semantics of navigational patterns and underlying user profiles. This research presents a novel method for online personalisation that incorporates data from user profiles and navigational behaviors. First, the information included in a collection of dynamic navigational patterns, a set of client profiles, and the anticipated time period were all reviewed. By assessing how similar the client profiles and navigational patterns were, the pages were then anticipated. The most important After that, the rank for each page was determined using user profiles and browsing habits. Lastly, it was suggested that online consumers browse the top n-pages with the highest rank. The most important Then, to determine the rank for each page, client profiles and navigational patterns were employed. Online consumers were ultimately cautioned against and given recommendations for the top n-pages with the highest ratings. The findings reveal that the suggested method successfully predicts the pages that web users would recommend.

References

[1] R. Baraglia, F. Silvestri, *An Online Recommender System for Large Web Sites*, in Proc. of ACM/IEEE Web Intelligence Conference (WI'04), China, 2004.

[2] J. Kleinberg, M. Sandler, *Using Mixture Models for Collaborative Filtering*, in Proc. of ACM Symposium on Theory of Computing (STOC'04), 2004

[3] O. Nasraoui, M. Pavuluri, Complete this Puzzle: A Connectionist Approach to Accurate Web Recommendations based on a Committee of Predictors, in Proc. of the 6th WEBKDD Workshop, Seattle, 2004.

[4] S. Acharyya, J. Ghosh, *Context-Sensitive Modeling of Web Surfing Behaviour Using Concept Trees*, in Proc. of the 5th WEBKDD Workshop, Washington DC, 2003

[5] J. Guo, V. Keselj, Q. Gao, Integrating Web Content Clustering into Web Log Association Rule Mining, In Proc. of Canadian AI, 2005

[6] X. Jin, Y. Zhou, B. Mobasher, *A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features*, in Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'05), Chicago, 2005

[7] Z. Huang, X. Li, H. Chen, *Link Prediction Approach to Collaborative Filtering*, in Proc. of ACM JCDL'05, 2005

[8] J. Borges, M. Levene, *Ranking Pages by Topology and Popularity within Web Sites*, accepted for publication in World Wide Web Journal, 2006

[9] D. Oberle, B. Berendt, A. Hotho, J. Gonzalez, Conceptual User Tracking, in Proc. of the 1st Atlantic Web Intelligence Conf. (AWIC), 2003

[10] H. Dai, B. Mobasher, *Using Ontologies to Discover Domain-Level Web Usage Profiles*, in Proc. of the 2nd Workshop on Semantic Web Mining, Helsinki, Finland, 2002

[11] Carsten Eickhoff, Kevyn Collins-Thompson, Paul Bennett, and Susan Dumais, Designing Human-Readable User Profiles for Search Evaluation, research.microsoft.com, 2014

[12] Apoorva A. Andhare, Nikita V. Mahajan, *Personalization of Web Knowledge Using Ontology Model*, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 2, pg. 348-353, 2014

[13] Akinori Nakamura, Nobuhiko Nishio, *User profile generation reflecting user's temporal preference through web life-log*, Proceedings of the ACM Conference on Ubiquitous Computing, Pages 615-616, 2012

[14] Zeynab fazelipour, Ali Harounabadi, *Personalization Web Pages for Site Users, Utilizing Users' Interests and Sequential Patterns Discovery*, Advances in Computer Science : an International Journal, Volume 4, Issue 6, No. 18, 2015

[15] Kuldeep Singh Rathore, Sanjiv Sharma, *Web Personalization Based on Enhanced Web Access Pattern using Sequential Pattern Mining*, International Journal Of Engineering And Computer Science, Volume 5, Issue 7, , pp. 17152-17159, 2016

[16] Dwarkanath Pande, S. ., & Hasane Ahammad, D. S. .

(2022). Cognitive Computing-Based Network Access Control System in Secure Physical Layer. Research Journal of Computer Systems and Engineering, 3(1), 14–20. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/36

[17] Kwame Boateng, Machine Learning in Cybersecurity: Intrusion Detection and Threat Analysis , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[18] Dhabliya, D. Security analysis of password schemes using virtual environment (2019) International Journal of Advanced Science and Technology, 28 (20), pp. 1334-1339.