

Web User Profiling through Quiescent Semantic Analysis for Web Personalization

Doddegowda B J¹, Sunil Kumar S Manvi², G T Raju³

¹ Research Scholar, Reva University, Asso. Prof. Dept. of CSE, AMCEC, Karnataka, India.

² Sunil Kumar S Manvi, Director, School of Computing & It, Reva University, Karnataka, India.

³ G T Raju, Professor, Dept. of CSE, RNSIT, VTU, Karnataka, India.

¹ bjdgowda10@gmail.com

² sunil.manvi@revainstitution.org

³ gtraju1990@yahoo.com

Abstract — Web personalization endorses the tailored Web pages or forecasts the personalized Web matters to Web users as per their particular observations. Fundamentally, user profile is generated for demonstrating particular user directional patterns derived from Web usage mining. This is done by corresponding the present active user session with the learned usage patterns. In this paper, a novel user profiling algorithm that makes use of usage patterns and a combined filtering approach is proposed. Firstly, similarity measure for Web user sessions is defined using cosine function of the user session over the quiescent semantic space. Then, the user sessions that exhibit similar navigational behaviors are clustered. The user profiles are viewed as the centroid of learned user session clusters. Whenever there is a fresh user session, the most corresponding user profile is chosen by determining distance between them. Then the top-N pages with the top support values are recommended. Experiments have been conducted on two data sets obtained from KDDCUP and RNSIT websites. The results show that the user profiles generated using the proposed method gives significant information for predicting and recommending the customized web pages for Web user.

Keyword - Web Personalization, User Profile, Quiescent Semantic Analysis, Singular Value Decomposition

I. INTRODUCTION

It is notable that Internet has turned out to be capable stage to store, spread and recover data. In any case, Web clients dependably experience the ill effects of the issues of data over-burden and suffocating because of huge and fast development in the measure of data and the quantity of clients. Due to this, issues like slight exactness and recall rate are two noteworthy worries that clients need to manage while looking for required data on Internet. Then again, the tremendous measure of information/data dwelling over the Internet contains a lot of important useful learning that could be found by means of cutting edge information mining approaches. Web personalization is a procedure that uses the instructive information gained from Web mining as a learning base, then predicts client potential get to inclinations, and prescribes the modified Web substance by alluding to the information base. This could be content, association, utilization and semantic data.

As on today, there are two sorts of methodologies usually utilized as a part of recommender frameworks, in particular content based filtering and collaborative filtering frameworks [1, 2]. Content based frameworks like Web Watcher [3] and customer side operator Letizia [4] normally create proposal in light of pre-developed client profiles by computing the likeness of Web substance to these profiles, while collaborative frameworks make suggestion by denoting other clients' inclination that is firmly like current one.

Computational cost is the key issue in clustering due to Web data features such as high-dimension and sparsity nature. *Quiescent Semantic Analysis (QSA)* is one of the effective dimensionality reduction algorithms explored here in order to address the issue. This method has the ability to realize the unseen information from Web data having semantic property into concern. *Quiescent Semantic Indexing (QSI)*, is a factual strategy, which is to reproduce a co-event perception space into a measurement lessened quiet space that keeps the greatest estimation of the first space by utilizing scientific change systems, for example, Singular Value Decomposition (SVD). With the lessened dimensionality of changed information expression, the computational cost is altogether diminished as needs be, and the issue of sparsity of information is taken care of also. Additionally, QSI based strategies are fit for catching the semantic learning from the perception information, while the ordinary factual examination methodologies, for example, grouping or arrangement are in absence of discovering fundamental relationship among the watched co-event. These days, QSI is widely embraced in

applications, for example, data recovery, picture handling, information mining and web mining. In this paper, we mean to coordinate QSI investigation with Web grouping procedures, to find Web client session totals with better quality.

The remainder of this paper is structured as: Section 2 portrays the hypothetical foundation of QSI strategy which depends on SVD estimation. Some essential ideas and recipes are exhibited to depict the points of interest of the calculation. Web grouping procedure for building client profiles is proposed by fusing peaceful semantic investigation in section 3. Trial results are provided at section 4. Related work is presented in section 5 and section 6 finishes up with an examination on scope for further work.

II. QUIESCENT SEMANTIC INDEXING METHOD

Here, the objective is to introduce QSI procedure and its connected fundamentals in mathematics, particularly the SVD, which is the foundation of QSI procedure. A novel function for similarity measure is then defined on the transformed semantic space in order to measure the distance between the user sessions.

A. Web Usage Data Model

Using the two sets User-sessions set $S = \{s_1, s_2, s_3, \dots, s_m\}$ and Web pages set $P = \{p_1, p_2, p_3, \dots, p_n\}$, we define usage-matrix where in the user-session is represented as ordered pairs of pages and their corresponding frequency of access. That is $s_i = \{(p_1, f_{i1}), (p_2, f_{i2}), \dots, (p_n, f_{in})\}$. In simple, each session can be written as a order of frequencies on the page space, that is $s_i = \{f_{i1}, f_{i2}, \dots, f_{in}\}$ where f_{ij} signifies the support for page p_j in user-session s_i . Consequently, the entire user-session data is represented as session-page matrix $SP\ m \times n = \{f_{ij}\}$ as shown in Table I.

TABLE I: USAGE-MATRIX

| Sessions/Pages | P1 | P2 | | Pj | |
|----------------|----|----|--|----|--|
| S1 | | | | | |
| S2 | | | | | |
| Si | | | | | |
| | | | | | |

In Table I, f_{ij} represents the frequency of access or the total time spent on corresponding page. Once the usage-matrix is built, we apply clustering procedure for gathering client sessions into different clusters. It is natural to perform grouping specifically on every column vector of the usage-matrix to decide the relative near session cluster by utilizing a similitude measure. Be that as it may, this sort of clustering system just catches the common connections between session information unequivocally. It is inept of uncovering the more profound basic qualities of usage-patterns. In this work, we propose Quiescent Usage Information (QUI) approach to deal with gathering client sessions semantically by considering the peaceful semantic data. For better comprehension QUI calculation, we first talk about some hypothetical foundations of the SVD technique.

B. SVD Method

The SVD description of a matrix is explained here. Consider $A = [a_{ij}]_{m \times n}$, real matrix without generalization loss and assuming $m \geq n$, there exists a SVD of A such that

$$A = U_{m \times m} \Sigma_{m \times n} V_{n \times n}$$

Where U and V are orthogonal matrices denoted as $U_{m \times m} = [u_1, u_2, u_3, \dots, u_m]_{m \times m}$ and $V_{n \times n} = [v_1, v_2, v_3, \dots, v_n]_{n \times n}$ where u_i ($i=1, 2, \dots, m$) is a m-dimensional vector $u_i = [u_{i1}, u_{i2}, u_{i3}, \dots, u_{im}]^T$ and v_j ($j=1, 2, \dots, n$) is another n-dimensional vector $v_j = [v_{j1}, v_{j2}, v_{j3}, \dots, v_{jn}]^T$. Suppose $rank(A) = r$ and singular values of A are the diagonal elements of Σ as follows:

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r \geq \sigma_r \geq \sigma_{r+1} \geq \dots = \sigma_n = 0$$

For a given threshold ϵ ($0 < \epsilon < 1$), parameter k is chosen such that $(\sigma_k - \sigma_{k+1}) / \sigma_k \leq \epsilon$

Then we denote $U_k = [u_1, u_2, u_3, \dots, u_k]_{m \times k}$, $V_k = [v_1, v_2, v_3, \dots, v_k]_{n \times k}$, $\Sigma_k = diagonal(\sigma_1, \sigma_2, \dots, \sigma_k)$ and $A_k = U_k \Sigma_k V_k^T$. It is well-known that A_k expresses the quiescent semantic information between the usage-data allows to find out relatively neighboring user-sessions at the semantic quiescent level based on their mutual similarity.

C. Illustration of User-Sessions in Quiescent Semantic Space

From the approximation matrix A_k , we re-write the user-sessions by mapping them into another k -dimensional quiescent semantic space. For example, session s_i re-written as $\{a_{i1}, a_{i2}, \dots, a_{ik}\}$, a coordinate vector with respect to pages. The projection of s_i in the k -dimensional quiescent semantic subspace is re-parameterized as

$$S_i = s_i V_k \sum^k = (t_{i1}, t_{i2}, \dots, t_{ik})$$

$$\text{Where } t_{ij} = \sum_{k=1}^n a_{ik} v_{kj} \sigma_j, \quad j = 1, 2, \dots, k$$

In this work, we assume the cosine function to identify mutual interests common in user-sessions. For example, the similarity between the two vectors $x=(x_1, x_2, \dots, x_k)$ and $y=(y_1, y_2, \dots, y_k)$ in a k -dimensional space is defined as

$$\text{Similarity}(x, y) = (x \cdot y) / (\|x\|_2 \|y\|_2)$$

$$\text{where } x \cdot y = \sum_{i=1}^k x_i y_i, \quad \|x\|_2 = \sqrt{\sum_{i=1}^k x_i^2}$$

Like this, we define the similarity between two user-sessions as:

$$\text{similarity}(s'_i, s'_j) = (s'_i \cdot s'_j) / (\|s'_i\|_2 \|s'_j\|_2)$$

III QUIESCENT SEMANTIC INDEXING METHOD

Here, we present an algorithm called *Quiescent Usage Information* (QUI) which first cluster the Web user sessions and next generate user profiles based on the centroids of the revealed clusters.

A. Procedure for Clustering User Sessions

We use modified *k-means* clustering algorithm to group user-sessions based on the transformed usage-matrix over the quiescent k -dimensional space. The parameter k needs not to be predefined. The algorithm is given below:

Algorithm 1: *Modified K-means clustering*

Input : session-page matrix SP and a similarity threshold ϵ

Output: set of user session clusters $USC = \{USC_i\}$ and the corresponding centroids $C_i = \{C_i\}$

Steps:

1. Choose the first user session S'_1 as the initial cluster USC_1 and C_1 , the centroid of this cluster.
2. For each session S'_i , calculate the similarity between S'_i and the centroids of other existing clusters \rightarrow similarity (S'_i, C_j) .
3. If similarity $(S'_i, C_k) = \max(\text{similarity}(S'_i, C_j)) > \epsilon$, then allocate S'_i into USC_k and recalculate the centroid of the cluster USC_k as $C_k = \sum_{j \in C_k} S'_j / |C_k|$
4. Otherwise, let S'_i form a new cluster and C_j be the centroid of this cluster
5. Repeat step 2 to 4 until all user sessions are processed and no more changes in centroids

B. Generating User Profile

As we specified over, every client session is spoken to as a support-based page vector. Along these lines, it is sensible to infer the centroid of the cluster acquired by the portrayed grouping procedure as user profile. Here, we figure the mean vector to speak to the centroid. For every client session group $USC_i \in USC$, the mean page vector of all sessions in the group is dictated by proportion of the entirety of page support in USC_i to the quantity of sessions in the group. With a specific end goal to take out the effect of contrast in invested energy or support of every session, the frequencies are standardized while figuring the centroid of group. That is, the most extreme support in the developed client profile is tuned to be 1, while other page supports are separated by the greatest support as needs be. In the meantime, some less-contributed pages (i.e. those with mean supports being short of what one certain cutoff) are sifted through. The procedure for developing client profile is given beneath:

Algorithm 2: *User Profile generation based on QSI*

Input : A set of user session clusters $USC = \{USC_k\}$

Output: A set of user profiles $UP = \{up_k\}$

Steps:

1. For each page p in the cluster USC_k , compute the mean support value of pages

$$ms(p, USC_k) = \sum_{s \in USC_k} S(p, s) / |USC_k|$$

Where $S(p,s)$ is the support of the page p in the session $s \in USC_k$ and $|USC_k|$ denotes the number of sessions in the cluster USC_k .

2. For each cluster, also calculate its mean vector (i.e. centroid) as $mv_c = \{ \langle p, ms(p, USC_k) \rangle \mid p \in P \}$
3. For each page within the cluster, if the value is less than the threshold μ , filter out the corresponding page, otherwise retain the page.
4. Sort pages with their support in a descending order and output the mean vector as the user profile.
 $up_k = \{ \langle p_{1k}, ms(p_{1k}, USC_k) \rangle, \langle p_{2k}, ms(p_{2k}, USC_k) \rangle, \dots \dots \langle p_{tk}, ms(p_{tk}, USC_k) \rangle \}$
 Where $ms(p_{1k}, USC_k) > ms(p_{2k}, USC_k) > \dots > ms(p_{tk}, USC_k) > \mu$
5. Repeat step 1 to 4 until all user session clusters are processed.

III. RESULTS AND DISCUSSIONS

Experiments have been conducted on two real world data sets obtained from log files of KDDCUP(www.ecn.purdue.edu/kddcup/) and RNSIT(rnsit.ac.in) web sites to evaluate the effectiveness of the proposed QUI algorithm. KDDCUP consists of 69 pages and 6305 user sessions after pre-processing. RNSIT consists of 48 pages and 5043 sessions after pre-processing. These data sets are articulated as usage-matrices with each column as page and each row as session. We use these matrices as an input data source and extract usage information and quiescent semantic associations. Then we apply the proposed algorithms to make Web recommendations. Table II shows the details of Page Ids, Pages and their frequency of access for the First Week of June 2013.

TABLE II. PAGE IDS, PAGES AND THEIR FREQUENCY OF ACCESS - ACADEMIC WEBSITE (RNSIT.AC.IN), JUNE 2013, FIRST WEEK

| Page Id | Web Page | Frequency Of Access(f) | Page Id | Web page | Frequency Of Access(f) |
|---------|----------------------|------------------------|---------|-----------------------|------------------------|
| 1 | /index.html | 26 | 25 | /Notifications.html | 0 |
| 2 | /Alumni.html | 80 | 26 | /Advertisements.html | 0 |
| 3 | /Admissions.html | 68 | 27 | /Invitations.html | 28 |
| 4 | /Departments.html | 40 | 28 | /awards.html | 5 |
| 5 | /library.html | 82 | 29 | /accolades.html | 6 |
| 6 | /Placement.html | 73 | 30 | /bestteachers.html | 25 |
| 7 | /hostel.html | 15 | 31 | /bestteachers-08.html | 2 |
| 8 | /Transportation.html | 14 | 32 | /mandatory.html | 9 |
| 9 | /Sports&Culture.html | 16 | 33 | /profile.html | 0 |
| 10 | /NSS.html | 16 | 34 | /Training.html | 2 |
| 11 | /contact.html | 69 | 35 | /Comp-visited.html | 25 |
| 12 | /courses.html | 75 | 36 | /mandatory-mech.html | 0 |
| 13 | /facilities.html | 23 | 37 | /mandatory-civil.html | 2 |
| 14 | /campus.html | 0 | 38 | /mandatory-ece.html | 3 |
| 15 | /pp_che.html | 15 | 39 | /mandatory-ee.html | 0 |
| 16 | /pp_maths.html | 0 | 40 | /mandatory-it.html | 2 |
| 17 | /pp_civil.html | 0 | 41 | /mandatory-CSE.html | 0 |
| 18 | /pp_ece.html | 7 | 42 | /mandatory-ISE.html | 2 |
| 19 | /pp_ee.html | 3 | 43 | /mandatory-mba.html | 6 |
| 20 | /pp_IT.html | 3 | 44 | /mandatory-mca.html | 2 |
| 21 | /pp_cse.html | 26 | 45 | /rank-holders.html | 22 |
| 22 | /pp_ise.html | 0 | 46 | /aboutus.html | 56 |
| 23 | /pp_mba.html | 3 | 47 | /vision&mission.html | 4 |
| 24 | /pp_mec.html | 3 | 48 | /Governing-body.html | 39 |

Sample user sessions for an academic website: rnsit.ac.in is shown in Table III. Similarly, Table IV presents sample user sessions & the corresponding page sequences and Table V gives the sample Session-Page matrix for Academic web site of 50 pages.

TABLE III. SAMPLE USER SESSIONS - ACADEMIC WEB SITE: RNSIT.AC.IN

| Session Id | IP Address | Date & Time | Page Accessed |
|------------|------------------|---------------------|--|
| 1 | 117.216.148.95 | 2014-04-30 17:57:03 | http://www.rnsit.ac.in/Invitations.html |
| | 117.216.148.95 | 2014-04-30 17:58:20 | http://www.rnsit.ac.in/facilities.html |
| 2 | 70.39.187.99 | 2014-04-30 18:03:38 | http://www.rnsit.ac.in/hostel.html |
| | 70.39.187.99 | 2014-04-30 18:03:39 | http://www.rnsit.ac.in /hostel.html |
| 3 | 210.212.194.2166 | 2014-04-30 18:27:52 | http://www.rnsit.ac.in /rank_holders.html |
| | 210.212.194.2166 | 2014-04-30 18:27:53 | http://www.rnsit.ac.in / rank_holders.html |
| 4 | 115.118.176.83 | 2014-04-30 18:28:42 | http://www.rnsit.ac.in /Admissions.html |
| | 115.118.176.83 | 2014-04-30 18:28:43 | http://www.rnsit.ac.in /Admissions.html |
| 5 | 117.208.191.46 | 2014-04-30 18:36:50 | http://www.rnsit.ac.in /cse-dep.html |
| | 117.208.191.46 | 2014-04-30 18:36:51 | http://www.rnsit.ac.in /cse-dep.html |

TABLE IV. SAMPLE USER SESSIONS AND THE CORRESPONDING PAGE SEQUENCES FOR ACADEMIC WEB SITE (RNSIT.AC.IN) OF 50 PAGES

| Session | uid | pids |
|---------|-----|-----------------|
| 1 | U5 | P1,P9,P24 |
| 2 | U23 | P32,P24,P5 |
| 3 | U34 | P36,P14 |
| 4 | U36 | P24,P14 |
| 5 | U4 | P46,P14,P27 |
| 6 | U27 | P9,P48,P24 |
| 7 | U17 | p2,P16 |
| 8 | U1 | P31,P26,P22,P12 |
| 9 | U2 | P1,P9,P14 |
| 10 | U35 | P20,P47 |

TABLE V. SAMPLE SESSION-PAGE MATRIX FOR ACADEMIC WEB SITE OF 50 PAGES

| Sessions /Pages | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P12 | P14 | P22 | P24 | P26 | .. | P31 | P32 | .. | P36 | |
|-----------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|----|-----|-----|----|-----|--|
| 1 – U5 | 26 | | | | | | | | 16 | | | | | 3 | | | | | | | |
| 2 – U23 | | | | | 82 | | | | | | | | | 3 | | | | 9 | | | |
| 4 – U36 | | | | | | | | | | | | 0 | | 3 | | | | | | | |
| 8 – U1 | | | | | | | | | | | 75 | | 0 | | 0 | | 2 | | | | |
| 9 – U2 | 26 | | | | | | | | 16 | | | 0 | | | | | | | | | |

Results after applying QUI algorithm on the two datasets are shown in Table VI and Table VII. In these tables, each user profile is represented by a sequence of important pages together with corresponding support expressed in a normalized form. Table VI depicts two user profiles generated from KDD dataset using QUI approach. The principal profile in Table VI speaks to the exercises required in web based shopping practices, for example, login, shopping basket, and checkout operation and so on., particularly happened in obtaining leg-wear items, while the second client profile mirrors the client's concern with respect to the retail chain itself.

TABLE VI. SAMPLE USER-PROFILES BUILT FROM KDDCUP DATASET

| Profile # | Page # | Page Title | Support |
|-------------|--------|--|---------|
| Profile - 1 | 29 | /Main-shopping_cart.html | 1.00 |
| | 4 | /Products-productDetaillegwear.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-dept_about.html | 0.34 |
| | 13 | /Articles-dept_about_mgmtteam.html | 0.15 |
| | 14 | /Articles-dept_about_boardofdirectors.html | 0.11 |

TABLE VII. SAMPLE USER-PROFILES BUILT FROM RNSIT DATASET

| 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 |

Similarly from Table VII, three profiles are generated: the first one reflects the students concerning issues regarding applying for admissions and the second one encompasses on facilities such as hostel, library, campus etc., third one indicates the students knowing more about computer science related courses and placements. We can observe from the generated profiles that most of the profiles show specific navigational behaviour whereas few represent multiple interests.

IV RELATED WORK

There are essentially two sorts of grouping methods utilized on web utilization information: Web exchange clustering and Web page grouping [12]. One utilization of Web page grouping is the versatile Web webpage. For example, PageGather [5, 6] has been proposed to orchestrate file pages that don't exist at first in view of dividing Web pages into different gatherings. The created list pages speak to the different get to interests of clients as per their navigational histories. Another case is that grouping client rating comes about has been embraced in cooperative sifting applications as an information get ready stride to enhance the adaptability of proposal utilizing k-Nearest Neighbor (kNN) strategy [7]. Mobasher et al. [11] use Web exchange and page bunching strategies to portray client get to designs for Web personalization in light of Web utilization information. These proposed grouping based methods have been turned out to be productive from their test comes about since they are truly equipped for recognizing the inborn normal traits uncovered from their memorable clickstream information. For the most part, these utilization examples are unequivocally caught at the level of client session or page. They, be that as it may, don't uncover the hidden attributes of client navigational exercises and additionally Web pages. For instance, such found use designs give little data of why such Web exchanges or Web pages are assembled together, and inactive connections among the co-event

perception information have not been consolidated into the mining forms also In [13], a calculation in light of Principal Factor Analysis (PFA) show got from factual examination, is proposed to produce client get to designs and reveal idle elements by bunching client exchanges and dissecting central variables required in the Web utilization mining. Proportionately, considers in [8-10] tended to determine client get to examples and Web page portions from different sorts of Web information, by using a Probabilistic Semantic Latent Analysis (PLSA) demonstrate, which depends on a most extreme probability rule from measurements. Amato et al. [14] utilizes topical client demonstrating for substance determination in advanced libraries. Their profiles concentrate on clients' inclinations in various spaces, for example, archive substance or structure. Nanas et al. [15] proposed a progressive profile in view of terms removed from clicked archives. In [16], the writers displayed and investigated the handiness of succinct comprehensible client profiles with a specific end goal to improve framework tuning and assessment by method for client considers. Survey on few customized models for web data gathering has been given in [17] and correlation has been made amongst them and a framework in light of personalization of web information utilizing philosophy is presented. The creators in [18] made utilization of the two qualities for finishing up client's transient necessity. To start with is to adjust the level of client's every enthusiasm with time run movement. Second, web perusing logs identified with an action. Despite the fact that the scientists contributed considerably in this field, yet they have not profoundly investigated how to produce smaller, comprehensible client profile representations. Hence, it is the need of great importance to create QSA-based methodologies that can uncover regular patterns unequivocally, as well as consider the quiet data verifiably amid mining.

V CONCLUSION

QSI based approach for clustering user sessions and creating user profiles has been proposed in this paper. Initially, the user sessions are modelled as usage-matrix and then SVD method has been used to apprehend the quiescent usage information for segregating user-sessions. A revised k-means clustering method which uses the QSI to create user session clusters also been proposed. These learned user-groups are then used to build user-profiles. The built user-profiles corresponding to various task oriented behaviours are signified as a set of page-support pairs, in which each support reflects the importance contributed by the page. Experiments have been conducted on two datasets to validate the effectiveness of the proposed QUI algorithm. The results shows that the proposed approach effectively discover the user profiles which reveal the essential relationships on the history of user interactions for the purpose of personalization. Future work in this regard is to integrate the information from user profiles and frequent sequential patterns for effective web personalization.

REFERENCES

- [1] Dunja, M., *Personal Web Watcher: Design and Implementation (Report)*. 1996, Department of Intelligent Systems, J. Stefan Institute, Slovenia.
- [2] Herlocker, J.L., et al., *Evaluating Collaborative Filtering Recommender Systems*. ACM Transaction on Information Systems (TOIS), 2004. 22(1): p. 5 – 53
- [3] Joachims, T., D. Freitag, and T. Mitchell. *Webwatcher: A Tour Guide For the World Wide Web*. in *The 15th International Joint Conference on Artificial Intelligence (IJCAI'97)*. 1997, p. 770-777, Nagoya, Japan.
- [4] Lieberman, H. *Letizia: An Agent that Assists Web Browsing*. in *Proc. of the 1995 International Joint Conference on Artificial Intelligence*. 1995, p. 924-929, Montreal, Canada: Morgan Kaufmann.
- [5] Perkowski, M. and O. Etzioni. *Adaptive Web Sites: Automatically Synthesizing Web Pages*. in *Proceedings of the 15th National Conference on Artificial Intelligence*. 1998, p. 727-732, Madison, WI: AAAI
- [6] Perkowski, M. and O. Etzioni. *Adaptive Web sites*. Communications of the ACM, 2000. 43(8): p. 152 - 158.
- [7] O'Conner, M. and J. Herlocker. *Clustering Items for Collaborative Filtering*. In *Proceedings of the ACM SIGIR Workshop on Recommender Systems*. 1999, Berkeley, CA, USA: ACM Press.
- [8] Xu, G., et al. *Discovering User Access Pattern Based on Probabilistic Latent Factor Model*. in *Proceeding of 16th Australasian Database Conference*. 2005, p. 27-36, Newcastle, Australia: ACS Inc.
- [9] Xu, G., Y. Zhang, and X. Zhou. *Using Probabilistic Semantic Latent Analysis for Web Page Grouping*. in *15th International Workshop on Research Issues on Data Engineering: Stream Data Mining and Applications (RIDE-SDMA'2005)*. 2005, p. 29-36, Tokyo, Japan
- [10] Jin, X., Y. Zhou, and B. Mobasher. *A Unified Approach to Personalization Based on Probabilistic Latent Semantic Models of Web Usage and Content*. In *Proceedings of the AAAI 2004 Workshop on Semantic Web Personalization (SWP'04)*. 2004, San Jose.
- [11] Mobasher, B., et al., *Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization*. Data Mining and Knowledge Discovery, 2002. 6(1): p. 61-82.
- [12] Mobasher, B., *Web Usage Mining and Personalization*, in *Practical Handbook of Internet Computing*, M.P. Singh, Editor. 2004, CRC Press. p. 15.1-37.
- [13] Zhou, Y., X. Jin, and B. Mobasher. *A Recommendation Model Based on Latent Principal Factors in Web Navigation Data*. in *Proceedings of the 3rd International Workshop on Web Dynamics*. 2004, New York: ACM Press.
- [14] G. Amato and U. Straccia. User profile modeling and applications to digital libraries. *Research and Advanced Technology for Digital Libraries*, 1999.
- [15] N. Nanas, V. Uren, and A. De Roeck. Building and applying a concept hierarchy representation of a user profile. In SIGIR 2003.
- [16] Carsten Eickhoff, Kevyn Collins-Thompson, Paul Bennett, and Susan Dumais, Designing Human-Readable User Profiles for Search Evaluation, *research.microsoft.com*, 2014.

- [17] 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.
- [18] 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.

AUTHOR PROFILE



Doddegowda B J, has received M.Tech Degree in Computer Science & Engineering from Visvesvaraya Technological University, Karnataka. Currently working as Associate Professor in the Department of Computer Science & Engineering, AMC Engineering College, Bengaluru, and pursuing Ph.D. Degree from Reva University, Bengaluru, Karnataka. He has over 14 years of experience in teaching. He has to his credit, publications of over 15 papers in National/International Conferences and Journals. His area of interests include Data Mining, KDD, Pattern Recognition.



Dr. Sunilkumar S. Manvi, has received his M.E. Degree from Bangalore University and Ph.D. Degree from IISc, Bengaluru. Currently working as Director, School of Computing and Information Technology, Reva University, Bengaluru, Karnataka. He has vast experience of more than 27 years in teaching in Electronics / Computer Science and Engineering. His research interests are: Software Agent based Network Management, Wireless Networks, Multimedia Networking Networks, Grid and Cloud computing, E-commerce. He has published around 150 papers in National Conferences, 100 papers in National and International Journals, and 15 Publications as books/book-chapters. He is a Fellow IETE, India, Fellow IE, India and SMIEEE, USA. He received best research publication award from VGST Karnataka in 2014. He has executed several research projects sponsored by government funding agencies.



Dr. G T Raju, has received M.E. Degree from Bangalore University, in 1995 and Ph.D. from Visvesvaraya Technological University, Karnataka in 2008. Currently working as Dean of Engineering, Professor and Head, Department of Computer Science & Engineering, RNS Institute of Technology, Bengaluru. He has 24 years of experience in teaching and research. His area of research interests include Web Mining, KDD, Image Processing, Pattern Recognition. He has published more than 90 papers in leading reputed International Journals/ Conference proceedings. He has authored five Technical books. He has completed two funded research projects. Ten Research Scholars have been awarded Ph.D. Degree under his supervision.