

Application of Incremental Relational Fuzzy Subtractive clustering Algorithm for Web Profile mining in Dynamic Environment

Aishwarya S¹, S.Vinoth Lakshmi^{*2}

¹ Assistant Professor, Department of Information Technology, Jerusalem College of Engineering, Chennai, India

^{2*} Assistant Professor, Department of Information Technology, Bharath University, Chennai, India

ABSTRACT: An improved understanding of the customer's habits, needs, and interests can allow the business to profit. Therefore reliable knowledge about the customers' preferences and needs forms the basis for effective Customer Relationship Management (CRM). As businesses move online, the competition between businesses to keep the loyalty of their old customers and to lure new customers is even more important, since a competitor's web site may be only one click away. In this paper, a complete framework and findings in mining web usage patterns from web log files of a real web site that has all the challenging aspects of real-life web usage mining, including evolving user profiles and external data describing an ontology of the web content. Even though the web site under study is part of a nonprofit organization that does not "sell" any products, it was crucial to understand "who" the users were, "what" they looked at, and "how their interests changed with time," all of which are important questions in CRM. In this paper an approach for discovering, tracking and evolving user profiles has been explored. Attempts have been made to describe how the discovered user profiles can be enriched with explicit information need that is inferred from search queries extracted from web log data. An objective validation strategy is also used to assess the quality of the mined profiles, in particular their adaptability in the face of evolving user behavior.

I. INTRODUCTION

Web usage mining is the application of data mining techniques to discover usage patterns from web [1-4]. Usage profiles give us information on the different ways in which the web site is being used. A useful application of this is web personalization, the process of automatically customizing the content and structure of a web site to the individual needs of each user [5]. In web mining most of the time the stress is on 'usage' profiles rather than 'user' profiles because it is not possible from server log data to distinguish individual users without explicit registration or client side tracking [6].

Profiles refer to information about individuals. Usage profiles capture different interests and trends among users accessing the site. These usage profiles can be used in personalization for recommendations, path prediction, perfecting of pages, better structuring of web sites, and in a nutshell improving the experience of the user. Usage profile information can also be used by content creators to understand which material is used more, how long a material is viewed by which types of users, and in what order material is accessed. The set of profiles reflects the semantics of the user historic dataset at particular point of time. Clustering is a widely used technique for discovering usage profiles from web log records.

Web personalization process can be divided into four distinct phases [10, 11], namely collection of web data, preprocessing of web data, analysis of web data, and recommendation. Recommender systems could be based on content-based filtering,

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

collaborative filtering, or rule-based filtering [11]. Moreover the clustering techniques, which find groups in data, are the most natural methods for generating usage profiles. Clustering can be done on object data or relational data. Relational data [12] provides a measure of the dissimilarity (or similarity) between objects and is represented by a matrix R , where R_{ij} , is dissimilarity between objects o_i and o_j . Also, $R_{ij}=0$, $R_{ij}=R_{ji}$ and $R_{ii}=0$. Relational data clustering is useful when object dimensions are non-numeric. Existing clustering techniques rely on some “carefully” specified input parameters such as number of clusters, initial partition, etc. Estimating values for these parameters is difficult. Another major problem with these clustering techniques is that they are compute intensive and do not scale for large data volumes like in web logs. Hence in this paper a relational fuzzy clustering technique is used and it does not require any user input, can handle noise and is scalable to large data sets.

II. STEPS INVOLVED IN DYNAMIC WEB MINING

The major steps involved in the web mining are listed down below

2.1. Collection of Web Log data and Database Updation

The Click streams and URLs that are navigated or visited by the user are updated continuously in the web log file. It is crucial to understand the different modes of usage and to know what kind of information the visitors seek and read on the web site and how this information evolves with time. The web log file is studied and data in the web log file is sorted and analyzed using special tokenize functionality and delimiter.

The analyzed data are taken as updated into the respective columns of the database which maintains the web log file in structured manner. This structured storage facility make it possible for us to implement the data mining algorithm to mine suitable rules and patterns based on which the user profile can be processed.

2.2 Preprocessing the Web Log File to Extract User Sessions

The clustering of the user sessions extracted from the web logs to partition the users into several homogeneous groups with similar activities and then extract user profiles from each cluster as a set of relevant URLs is called as preprocessing.

2.2.1. Adding Semantics with Ontology Concepts

When clustering the user sessions, it customary to exploit the Web site hierarchy to give partial weights in the session similarity between URLs that are distinct and yet located closer together on this hierarchy. The Web site hierarchy is inferred both from the URL address and from a Web site database that organizes most of the dynamic URLs along an “is-a” ontology of items. The cluster profiles are also enriched with various facets, including search queries submitted just before landing on the Web site, and inquiring and inquired companies, in case users from (inquiring) companies inquire about any of the (inquired) companies listed on the web site, which provide related services. Relying only on web usage data for user modeling or for personalization can be inefficient, either when there is insufficient usage data for the purpose of mining certain patterns or when new pages are added and thus do not accumulate sufficient usage data at first. The lack of usage data in these cases can be compensated by adding other information such as the content of web pages or the structure of a web site. In the keywords that appear in web pages are used to generate document vectors, which are later clustered in the document space to further augment user profiles. In the Web site’s own hierarchical structure is treated like an implicit taxonomy or concept hierarchy that is exploited in computing the similarity between any two web pages on the web site. This allows a better comparison between sessions that contain visits to web pages that are different and yet semantically related (for example, under the same more general topic). The idea of exploiting concept hierarchies or taxonomies has already been found to enhance association rule mining and to facilitate information searching in textual data.

2.2.2. Mapping of Content Keywords

A dynamic URL is a page address that results from the search of a database-driven Web site or a Web site that runs a script. Unlike static URLs, in which the contents of the Web page do not change, dynamic URLs are typically generated from

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

specific queries to a site's database. Even though the examples given in the following discussion consistently use the ASP extension, this extension can be replaced by any other dynamic URL extension (such as PHP), without any changes in generic approach. Although static Web pages tend to have meaningful URLs such as /reports/fall_2003/benefits.html, most dynamic URLs such as /universal.aspx?id=55&codes_id=60 are unfortunately hard to discern or even recognize based only on their URL. The issue is generally resolved by resorting to available external data that maps database contents to a dynamic resource and its parameter values. The ASP codes in most menus can be mapped during the preprocessing phase to a parent/child structure by using external data, thus mapping URLs to meaningful hierarchical descriptions.

2.3 Clustering of the user sessions by using Incremental Relational Fuzzy Subtractive Algorithm(IRFSC)

IRFSC Algorithm offer better performance compared to other clustering algorithm. This algorithm can handle noise in better way and automatically determines the number of clusters. In addition, evolutionary optimization allows the use of any domain-specific optimization criterion and any similarity measure, in particular a subjective measure that exploits domain knowledge or ontology's. The data is continuously pre-processed to produce session lists: A session list s_i for user i is an item list of URLs visited by the same user. In discovery mode, a session is fed to the learning system as soon as it is available. The i^{th} B-Cell, D-W-B-cell (i), represents the i^{th} candidate profile and encodes a list of relevant URLs. It is matched to incoming sessions using 1-Cosine similarity as a distance measure. Each profile has its own influence zone defined by sigma square (i). This measures the average dissimilarity between the i^{th} candidate profile and the sessions/antigens that activate the i^{th} B-Cell, D-W- B-cell (i).

2.4 Clusters/Categories the user session into user profiles

After automatically grouping sessions into different clusters, summarizing the session categories in terms of user profile vectors is done. The k^{th} component/weight of this vector (p_{ik}) captures the relevance of URL (k) in the i^{th} profile, as estimated by the conditional probability that URL_k is accessed in a session belonging to the i^{th} cluster (this is the frequency with which URL_k was accessed in the sessions belonging to the i^{th} cluster). The profiles are then converted to binary vectors (sets) so that only URLs with weights > 0.15 remain. The model is further extended to a robust profile based on robust weights (w_{ij}) computed in the algorithm that assign only sessions with high robust weights (that is, $w_{ij} > w_{\min}$) to a cluster's core. The core of a profile consists only of sessions that are very similar to the representative profile.

2.5 Track current profiles against existing profiles

Tracking different profile events across different time periods can generate a better understanding of the evolution of user access patterns and seasonality. Note that both profiles and click streams are typically evolving, since the profiles are nothing more than summaries of the click streams, which are themselves evolving. Each profile p_i is discovered along with an automatically determined measure of scale sigma (i) that represents the amount of variance or dispersion of the user sessions in a given cluster around the cluster representative. This measure is used to determine the boundary around each cluster (an area located at a distance i from the profile p_i) and thus allows us to automatically determine whether two profiles are compatible. Two profiles are compatible if their boundaries overlap. The notion of compatibility between profiles is essential for tracking evolving profiles. After mining the Web log of a given period, we perform an automated comparison between all the profiles discovered in the current batch and the profiles discovered in the previous batch by a sequence of SQL queries on the profiles that have been stored in a database, as shown in the "Track Profiles" Algorithm [13-14].

III. PROPOSED METHODOLOGY

3.1 System Flow Diagram

The step involved in analysis of web log is given as system flow cart in Fig 1.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

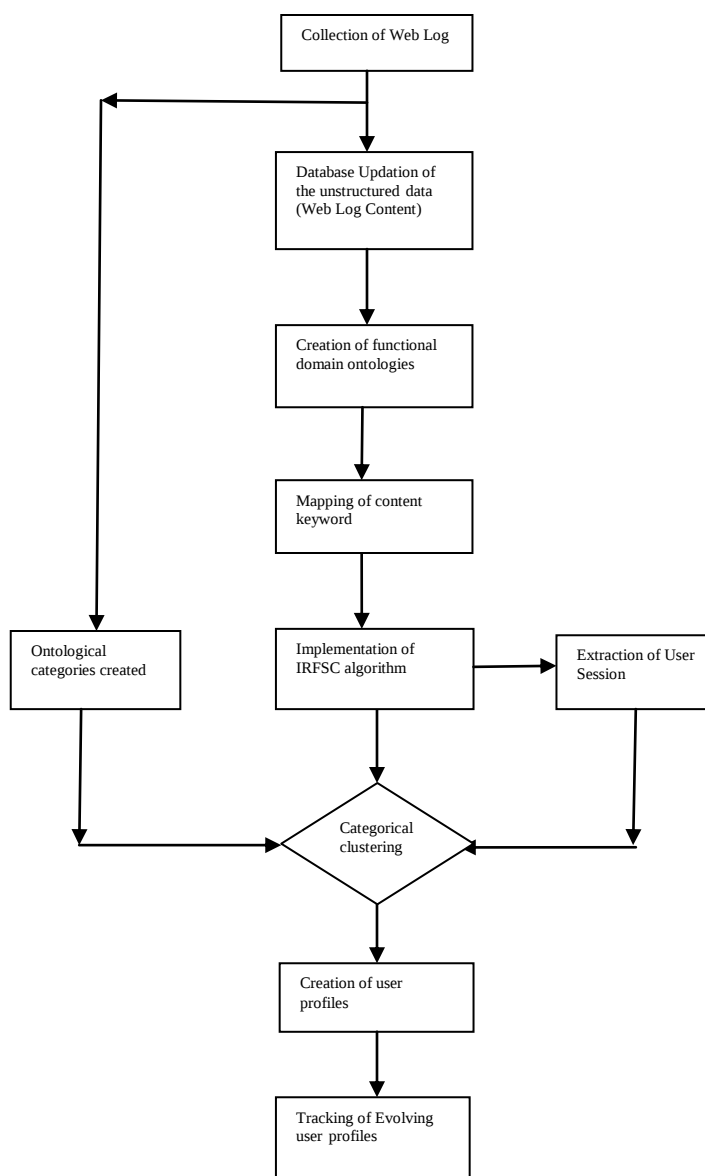


Figure 1: System Flow Diagram

The details related to web mining has been discussed in the earlier sections. It may broadly be summarized as collection of web log, clustering of data, and creation of user profile followed by tracking.[1]

Event state states the idleness of the CPU and in host availability; it could be the monitoring information of a particular host.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

3.2 IRFSC ALORITHM FOR CLUSTERING

Database contents or web usage data grow dynamically and it requires some cluster maintenance scheme by which these changing trends and patterns can be captured without having to apply frequently this relatively expensive operation of re-clustering of large volume of old and new data together. Cluster maintenance however enables us to adapt to the dynamic and changing environment in a much less expensive manner in terms of computation times and resources and also enables consecutive maintenance even after re-clustering. Thus an optimal combination of full data clustering and cluster maintenance is ideally suited for dynamic environments. In this section, we describe IRFSC algorithm that is an extension of the RFSC algorithm and forms part of our usage profile maintenance scheme.

```

1) Calculate the potential  $P_{new}$  of the object to be inserted  $x_{new}$ 
2) Raise the potentials of all existing cluster prototypes;
    $D_{i, new}$  = dissimilarity between  $x_{new}$  and  $i^{th}$  cluster center
   
$$P_i = P_i + e^{-\alpha D_{i, new}^2}$$

3)  $d_{min}$  = minimum of the dissimilarity values between  $x_{new}$  and all
   the previously found cluster centers.
   for  $i=1$  to  $C$  do
     if ( $P_{new} > P_i$ ) then
       if ( $P_{new} \geq \epsilon P_1$ ) //case (iii) true
          $I += d_{min}$ 
       else if ( $P_{new} \geq \epsilon P_1$ )
         If  $\left(\frac{d_{min}}{\gamma}\right) + \left(\frac{P_{new}}{P_1}\right) \geq 1$  then //case
         (iii) true
            $I += d_{min}$ 
         end if
       end if
     else
       
$$P_{new} = P_{new} + e^{-\alpha D_{i, new}^2}$$

       If ( $P_{new} < 0$ ) goto step 5; // no further comparisons required
       as case (i) is true
     end if
   end for
4) if (( $P_{new} > 0$ ) and (Case (iii) is not true))
   //check if  $x_{new}$  can become a new cluster center
   if ( $P_{new} \geq \epsilon P_1$ )
     If  $\left(\frac{d_{min}}{\gamma}\right) + \left(\frac{P_{new}}{P_1}\right) \geq 1$  then //case (ii) true
     //  $x_{new}$  is the new cluster center
      $C = C + 1$ ; //increment cluster count
      $P_C = P_{new}$ ;
     Calculate the membership of  $N_U$  old objects
     with respect to this new cluster;
     end if
   end if
5) Increment  $N_U$  by 1 as a new object is added;
   Calculate membership of  $x_{new}$  with  $C$  clusters.

```

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

3.3 Data Flow Diagram

The data flow diagrams at various stages are given in the flow chart below.

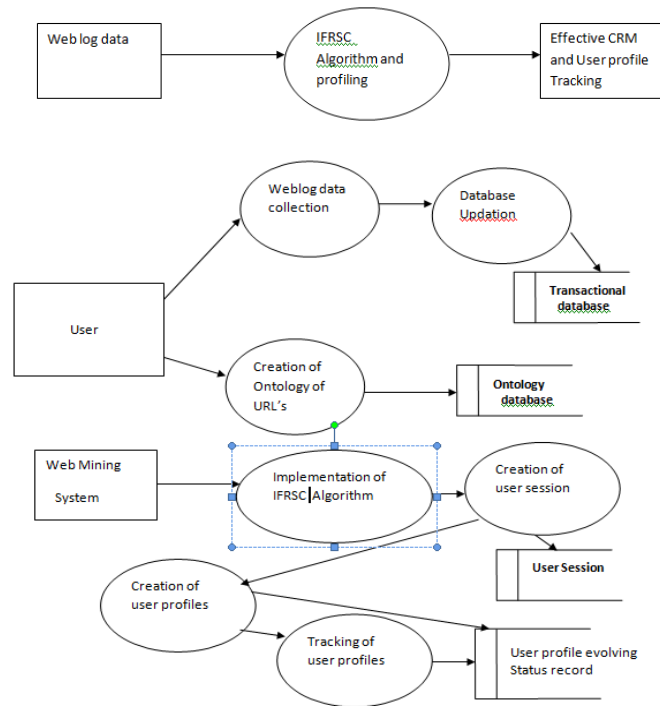


Figure 2: Data Flow Diagram in the present Work

The user weblog data is stored in the transactional database and the same time ontology of the URL's are also created and stored in the ontology database. The IFRSC algorithm is used to analyse the user session and followed by tracking of the user profiles

3.4 Cluster Maintenance using IFRSC

Database contents or web usage data grow dynamically. Hence, cluster maintenance has gained paramount importance.[2] Moreover some cluster maintenance scheme by which these changing trends and patterns can be captured without having to apply frequently this relatively expensive operation of re-clustering of large volume of old and new data together. It can therefore at best be a close approximation to clustering of the complete data, old plus new. Thus re-clustering will always give more accurate results and will be required to be done periodically. Cluster maintenance however enables us to adapt to the dynamic and changing environment in a much less expensive manner in terms of computation times and resources and also enables consecutive maintenance even after re-clustering.[3] Thus an optimal combination of full data clustering and cluster maintenance is ideally suited for dynamic environments. Let us assume that we have a set of N_u objects and C clusters. Also $u(C \times N_u)$ is the membership matrix which contains the membership values of each object to each of the C clusters.[4] To begin with, these C clusters and the matrix u are obtained using RFSC algorithm. Subsequently, these are

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

updated using incremental RFSC until re-clustering of the complete dataset is required. These cycles of incremental RFSC and re-clustering make up the maintenance scheme. It can be explained as follows

Let X_{new} be the new object which is required to be inserted. Depending upon X_{new} we have three possible situations:

- Case i) X_{new} has a high membership value with respect to one of the existing clusters and hence its potential is such that it becomes a core member of that cluster or it has a very low potential in which case it is probably a noise.[5]
- Case ii) The potential of X_{new} is such that it is not a core member of any of the clusters, its potential is lower than all the potentials of the C clusters, but is high enough to become a new cluster center.
- Case iii) The potential of X_{new} is such that it is better than the potential of one of existing cluster centers. [6]

The process of determining the case that is applicable to the new object and the actions taken are described below:

1. First, the stored potentials of all previous cluster prototypes P_i are raised by the degree to which this X_{new} is close to these previously found cluster prototypes.[7] Hence the potential of the cluster prototype will increase by a larger degree if X_{new} is closer to it (as X_{new} makes that cluster more dense) than when X_{new} is farther away. The maximum increase in potential is 1, i.e., when X_{new} coincides with an existing cluster center. Let these updated potentials be P'_i .

2. If X_{new} has potential better than the first cluster center, i.e., $P_{new} > P'_1$. If yes, then Case (iii) holds and the actions taken are described later below[8]. If not, proceed with the subtractive step, i.e., potential of X_{new} will be reduced depending on its similarity with the first cluster center.[9]

After this subtractive step, if P_{new} becomes negative, it means that X_{new} is a core element or very close to the first cluster center (Case (i) is true). No further comparison or subtraction is required.[10]

3. If the potential does not become negative, then it can be compared with the updated potential of X_{new} with the potential of next cluster prototype. Again if the potential of X_{new} is less than this cluster center, then subtractive step is performed and the above discussion holds true. [11]

If the potential of X_{new} is greater than the current cluster center, i.e., X_{new} has potential better than the potential of an existing cluster center (Case (iii) holds) then continue with step 4a, otherwise proceed with the subtractive step.

Also it might so happen that after comparison with the potentials of all the existing C cluster prototypes and going through the corresponding subtractive steps X_{new} may have a potential good enough to become a new cluster center (Case (ii) is true) [13].

IV. EXPERIMENTAL RESULTS

The results of the user sessions updated for web mining is discussed in the following sections.[13]

4.1 Updating of the Web logs

In the present work user access logs from various time intervals in the month was accessed and the preprocessing step was carried out and created sessions by keeping the maximum elapsed time between two consecutive accesses from a particular IP address as 45 minutes.[15] We also removed short sessions of length 1 or 2, as they do not provide much useful information. Compared to a few hundred or thousand sessions used for verifying other fuzzy clustering techniques in our work, the number of user sessions as well as the dimensions is much larger. The screen shots for updating of the data is given in Fig 3.[16]

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

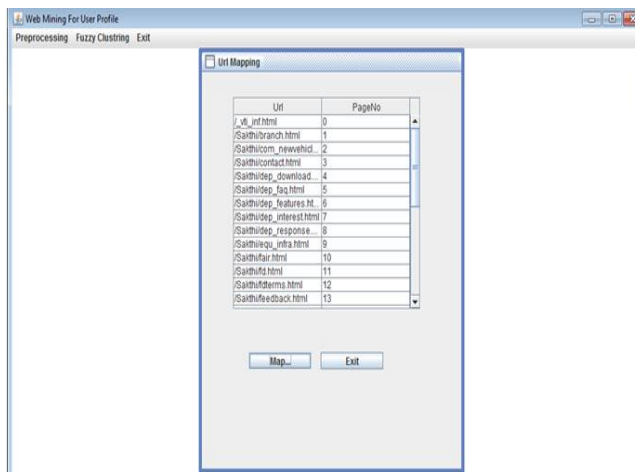


Figure 3: Screen Shot for updating of the web logs

To find the optimal values of ε and ε' , search for best clustering of the data was carried as follows. The optimal values were varied from 0.1 to 1 in steps of 0.1, with the condition that $\varepsilon < \varepsilon'$, apply the clustering algorithm for each pair of ε and ε' and choosing the one with minimum value for cluster goodness index. The updated weblog is mapped for different URL's and user control vectors are generated as given in the Fig 4[12].

4.2 IRFSC clustering

The clustering of the user profiles were done using IRFSC algorithm and the clustered profiles after execution is given in Fig 5 The usage profile e for the i th cluster is U_{Pi} , which is a tuple with NU components, $(up_{i1}, \dots, up_{iNU})$, where NU is total number of distinct URLs and up_{ij} is the number of sessions in cluster i in which the j th URL is clicked divided by the number of sessions in cluster i . Intuitively, up_{ij} indicates the popularity of URL j in cluster i . The clustered profiles are arrange according to the popularity as given in the Figure 5

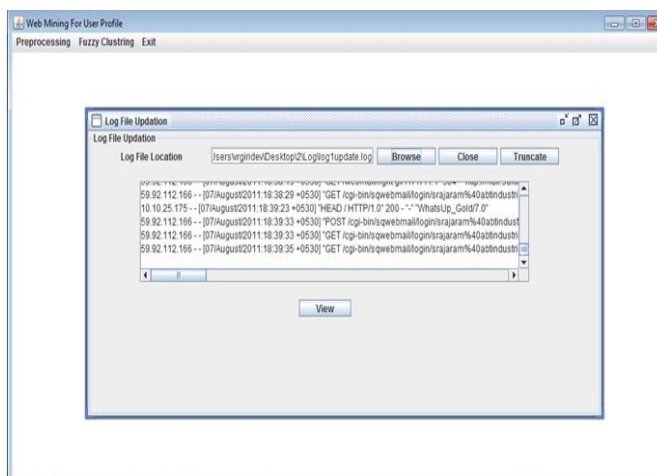
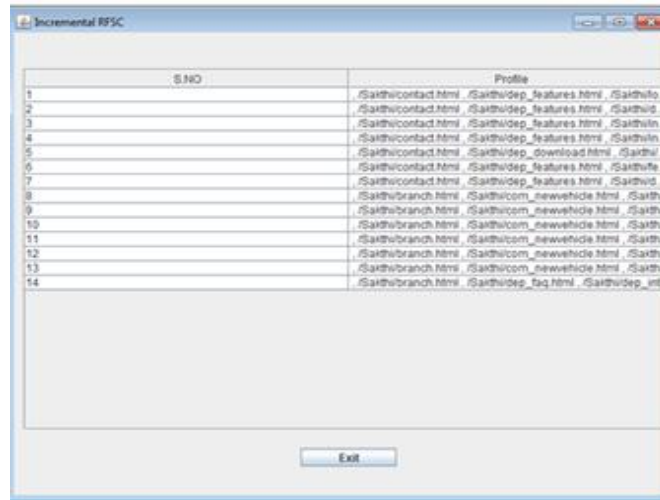


Figure 4: Url mapping for web log generated

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015



S.NO	Profile
1	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
2	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
3	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
4	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
5	/Sakthi/contact.html /Sakthidep_download.html /Sakthi/
6	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
7	/Sakthi/contact.html /Sakthidep_features.html /Sakthi/
8	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
9	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
10	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
11	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
12	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
13	/Sakthi/branch.html /Sakthi/com_newvehicle.html /Sakthi/
14	/Sakthi/branch.html /Sakthidep_faqs.html /Sakthidep_int

4.3 Profile Evolution events and its context with respect to marketing

The mass profile evolution events and its context in relation to marketing is given in Table 1

Table 1. Explanation of Tracking profiles from weblog

Event	Explanation	Implication in marketing
Birth	A new group of user clicks that has not been observed	Indicates emerging market segment
Death	User activity that vanishes	Decline phase of segment
Persistence	Observed in two continuity periods	Indicates sustainability of the market
Atavism	Profile temporally vanishes and reemerges	Indicates the product has regained interest

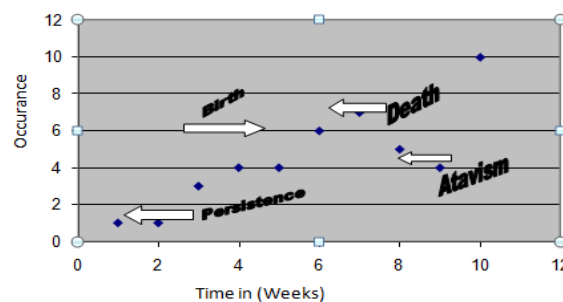


Figure 6: Web profile mining and its marketing implications

The emerging trends of the user profiles are given in Fig 6.[17] It can be seen that in our case there is persistence which is

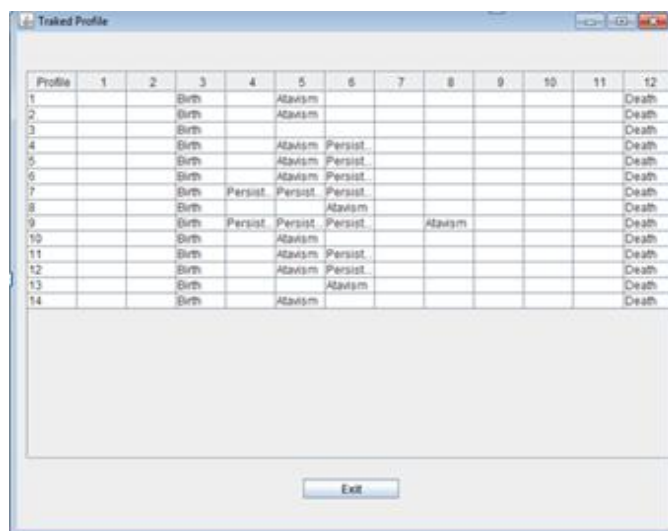
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

observed in the first two weeks followed by death where in user has lost interest in the intended web product and there is increased interest after six weeks followed by Atavism where in you sudden spurt in the occurrence of the web log indicating that the consumer has regained interest in the product

The analysis of profile evolution, as shown in figure above, can improve our understanding of the user activity trends and detect seasonality in their access patterns, especially over a long time span.[18] It also helps in implementing a dynamic recommendation strategy, for instance, by caching frequently reemerging (atavistic) profiles. Dead profiles can be relegated to the secondary memory for possible reemergence, whereas persistent profiles can be kept in the primary memory for fast access and then relegated to the secondary memory when they die. Similarly, dead profiles that have been persistent during an earlier period should be distinguished from dead profiles that have never been persistent, that is, volatile profiles.[19]



Profile	1	2	3	4	5	6	7	8	9	10	11	12
1			Birth		Atavism							Death
2			Birth		Atavism							Death
3			Birth									Death
4			Birth		Atavism	Persist						Death
5			Birth		Atavism	Persist						Death
6			Birth		Atavism	Persist						Death
7			Birth	Persist	Persist	Persist						Death
8			Birth			Atavism						Death
9			Birth	Persist	Persist	Persist		Atavism				Death
10			Birth		Atavism							Death
11			Birth		Atavism	Persist						Death
12			Birth		Atavism	Persist						Death
13			Birth			Atavism						Death
14			Birth		Atavism							Death

Figure 7: Tracking of user profiles

V. CONCLUSION

In this paper a framework for mining, tracking, and validating evolving multifaceted user profiles on Web sites that have dynamic environment has been carried. The present work provided scope for analysis of the user sessions and the valuable information regarding customer accesses pattern are analyzed. The analyzed data provides scope for customer related marketing. The period length for analysis depends on the application or can be set, depending on the cross-period validation results. The performance of the IRFSC algorithm is being compared with H-UNC Algorithm as a part of further work, which will give better idea of clustering of data in short span of time.

REFERENCES

1. R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," Proc. Ninth IEEE Int'l Conf. Tools with AI (ICTAI '97), pp. 558-567, 1997.
2. O. Nasraoui, R. Krishnapuram, and A. Joshi, "Mining Web Access Logs Using a Relational Clustering Algorithm Based on a RobustEstimator," Proc. Eighth Int'l World Wide Web Conf. (WWW '99), pp. 40-41, 1999.
3. Sree Latha R., Vijayaraj R., Azhagiya Singam E.R., Chitra K., Subramanian V., "3D-QSAR and Docking Studies on the HEPT Derivatives of HIV-1 Reverse Transcriptase", Chemical Biology and Drug Design, ISSN : 1747-0285, 78(3) (2011) pp.418-426.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2015

4. O. Nasraoui, R. Krishnapuram, H. Frigui, and A. Joshi, "Extracting Web User Profiles Using Relational Competitive Fuzzy Clustering," *Int'l J. Artificial Intelligence Tools*, vol. 9, no. 4, pp. 509-526, 2000.
5. J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "WebUsage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations*, vol. 1, no. 2, pp. 1-12, Jan.2000.
6. Masthan K.M.K., Aravindh Babu N., Dash K.C., Elumalai M., "Advanced diagnostic aids in oral cancer", *Asian Pacific Journal of Cancer Prevention*, ISSN: 1513-7368, 13(8) (2012) pp.3573-3576.
7. M. Spiliopoulou and L.C. Faulstich, "WUM: A Web Utilization Miner," *Proc. First Int'l Workshop Web and Databases (WebDB '98)*, 1998.
8. Breese, J., Heckerman, D., Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proc. of UAI-98*, pp. 43-52, 1998.
9. Tamilselvi N., Dhamotharan R., Krishnamoorthy P., Shivakumar, "Anatomical studies of *Indigofera aspalathoides* Vahl (Fabaceae)", *Journal of Chemical and Pharmaceutical Research*, ISSN : 0975 – 7384 , 3(2) (2011) pp.738-746.
10. Can, F., Ozkaran, E.A. A dynamic cluster maintenance system for information retrieval. In *Proc. of the 10th Annual International ACM-SIGIR Conference*, pp. 123-131, 1987.
11. Charikar, M., Chekuri, C., Feder, T., Motwani, R., 2004. Incremental clustering and dynamic information retrieval. In *SIAM Journal on Computing* 33 (6), pp. 1417-1440, 2004.
12. Devi M., Jeyanthi Rebecca L., Sumathy S., "Bactericidal activity of the lactic acid bacteria *Lactobacillus delbreuckii*", *Journal of Chemical and Pharmaceutical Research*, ISSN : 0975 – 7384 , 5(2) (2013) pp.176-180.
13. Cooley, R., Mobasher, B. and Srivastava, J. Data reparation for Mining World Wide Web Browsing Patterns. *Journal of Knowledge and Information Systems*, 1, pp. 1-27, 1999.
14. Ester, M., Kriegel, H., Sander, J., Wimmer, M., Xu, X. Incremental clustering for mining in a data warehousing environment. In *Proc. of VLDB 1998*, Morgan Kaufmann Publishers Inc., pp. 323-333, 1998.
15. Hathaway, R.J., Bezdek, J.C., Davenport, J.W. On relational data version of c-means algorithm. *Pattern Recognition Letters*, 17, pp. 607-612, 1996.
16. Hubert, L. & Arabie, P. Comparing partitions, *Journal of Classification*, 2, pp. 193-198, 1985.
17. Jain, A. K., Dubes, R. C. Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs, N. J., 1988.
18. Reddy Seshadri V., Suchitra M.M., Reddy Y.M., Reddy Prabhakar E., "Beneficial and detrimental actions of free radicals: A review", *Journal of Global Pharma Technology*, ISSN : 0975-8542, 2(5) (2010) pp.3-11.
19. Mobasher, B. Web Usage Mining and Personalization. *Practical Handbook of Internet Computing*, Munindar P.Singh (ed.), CRC Press, 2004.
20. B Karthik, TVUK Kumar, A Selvaraj, Test Data Compression Architecture for Lowpower VLSI Testing, *World Applied Sciences Journal* 29 (8), PP 1035-1038, 2014.
21. .M.Sundararajan .Lakshmi,"Biometric Security system using Face Recognition", *Publication of International Journal of Pattern Recognition and Research*. July 2009 pp. 125-134.
22. .M.Sundararajan," Optical Sensor Based Instrumentation for correlative analysis of Human ECG and Breathing Signal", *Publication of International Journal of Electronics Engineering Research*, Research India Publication, Volume 1 Number 4(2009). Pp 287-298.
23. .C.Lakshmi & Dr.M.Sundararajan, "The Chernoff Criterion Based Common Vector Method: A Novel Quadratic Subspace Classifier for Face Recognition" *Indian Research Review*, Vol.1, No.1, Dec, 2009.
24. .M.Sundararajan & P.Manikandan," Discrete wavelet features extractions for Iris recognition based biometric Security", *Publication of International Journal of Electronics Engineering Research*, Research India Publication, Volume 2 Number 2(2010).pp. 237-241.
25. .M.Sundararajan, C.Lakshmi & .M.Ponnaivaikko, "Improved kernel common vector method for face recognition varying in background conditions", *proceeding of Springer – LNCS 6026- pp.175-186 (2010).ISSN 0302-9743.(Ref. Jor – Anne-II)*