RESEARCHERID
THOMSON REUTERS

[COTII- 2018]

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

WEB USAGE MINING BASED PERSONALIZATION SYSTEM- A REVIEW AND IMPLEMENTATION

Dr.Bhawesh Kumar Thakur[*1] & Dr. Sayed Qamar Abbas[2]

[*1&2]Deptt. of Computer Science and Engineering, Ambalika Institute of Management & Technology, AKTU, Uttar Pradesh ,Lucknow, India

ABSTRACT

The explosive development in the number and the intricacy of data assets and availability of services on the Web, log based information become an authentic and genuine asset to describe the users for Web. It makes data of related web information as chain of importance structure. Such structure can be utilized as the contribution for an assortment of information mining activities such as clustering. It is useful for e-commerce application that uses clusters to make personalized environment for the Web users.

Personalization of client information is a current marvel because of the ascent of Internet utilization, yet finished the years Web mining procedures have built up an undeniable help to give customized understanding to client. Part of work is done to ad lib Web Personalization.

The paper provides an insight on various approaches used for providing a personalized feeling for users automatically when she/he associating with web by grouping of web utilization information utilizing idea chain of importance. This paper also present efficient personalization logic based on the web server's access logs by methods for information and web usage mining procedures.

Keywords: *Web Mining, Web Usage Mining, Web Usage Data, Personalization*

I. INTRODUCTION

Web mining refers to the use of data mining techniques to automatically retrieve, extract and evaluate (generalize/analyze) information for knowledge discovery from Web documents and services. Web data is typically unlabelled, distributed, heterogeneous, semi-structured, time varying, and high dimensional. Hence any human interface needs to handle context sensitive and imprecise queries, and provide for summarization, deduction, personalization and learning. Almost 90%of the data is useless, and often does not represent any relevant information that the user is looking for. Taking into account the huge amount of data storage and manipulation needed for example a simple query; the processing essentially requires adequate tools suitable for extracting only the relevant, sometimes hidden, knowledge as the final result of the problem under consideration [1].

Many Web sites are big and complicated and people generally deviate from the goal of given query, or obtain unpredicted outcomes when they attempt to explore using them. So, the e-commerce based websites are quickly evolving and the requirement for predicting the needs of the customers is more obvious. Therefore, user needs prediction can be used to improve the frequency of users and their retention on a Web site and it can only be managed by personalization.

Web Personalization is defined as an act that is comfortable with the facilities i.e. information or services provided by the web site and with the help of the knowledge gained from the users' navigational behavior and each and every user interests fulfill the requirements of a individual or a group of users.

The objective of a Web personalization is to [2] "provides users with the information they want, without expecting from them to ask for it explicitly".

ISSN 2348 – 8034
Impact Factor- 5.070

Web Usage Mining of the data generated by the users 'interactions with the Web, typically represented as Web server access logs, user profiles, user queries and mouse-clicks. This includes trend analysis (of the Web dynamics information), and Web access association/sequential pattern analysis. Web usage mining performs mining on Web usage data, or web logs. A web log is a listing of page reference data. Sometimes it is referred to as Click streams data because each entry corresponds to a mouse click.

These logs can be examined from either a client perspective or a server perspective. When evaluated from a server perspective, mining uncovers information about the sites where the server resides. It can be used to improve the design of the sites. By evaluating a client's sequence of clicks, information about a user or group of users is detected. This could be used to perform pre-fetching and caching of pages.

A. **Web Personalization**

Personalization aims to provide users with what they need without explicitly requiring them to request for it. This means that a personalization system must somehow infer what the user requires based on either previous or current interactions with the user (Anand, Sarabjot singh and Mobasher, Bomshad, 2005) [3]. In web mining and web personalization context, two terms are widely used interchangeably. So, it is necessary to explore the concept of customization and personalization. Customization process helps the site to adjust according to each user's preferences related to its structure and presentation. When a registered user logs in, customized home page is displayed to her/him. In personalization systems dynamic modifications of a Web site related to the web content or the site structure are executed.

Main components of Web personalization include modeling of Web objects and subjects i.e. users, categorization of objects and subjects, matching between and across objects and/or subjects, and identification of the set of actions to be recommended for personalization.
Personalization can be identified as a type of clustering. Web personalization is an important task from the user point of view as well as application point of view. Web personalization helps the organizations to develop customer-centric Web sites.

Main elements of Web personalization system includes:
a) the classification and preprocessing of web data,
b) finding association among several types of classified data, and
c) The identification of the recommendation which should be suggested by given personalization system.

B. **Clustering**

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Dissimilarities are assessed based on the attribute values describing the objects. Clustering has its roots in many areas including data mining, statistics, biology and machine learning.

The objects in data mining could have hundreds of attributes. Clustering in such high dimensional spaces presents remarkable difficulty, much more so than in predictive learning.

With the support of an effective method cluster can be constructed based on the concept hierarchy and α-similarity. The new model to cluster the web user transactions based on the concept hierarchy of web usage data and fuzzy proximity relations of user transactions was tested for its effectiveness.
The fuzzy concept ensemble to identify amount of similarities between us
er transaction clusters. The user specified α value is an important factor based on which the performance of cluster depends.

II. LITERATURE REVIEW

In last several years, many activities related to the research deals with the Web usage mining and Web personalization is conducted. Many of the work emphasizes on retrieving useful patterns and rules with the help of data mining approaches to recognize the users' access behavior, so the conclusion can then be taken related to website reformation or updating. The users navigate through a site and recommendation engine assists the user to navigate numerous times. Some of the further enhanced web based systems provide much more functionality providing way of dynamically modifying a site's structure.

Personalization of user data is a recent phenomenon due to the rise of Internet usage, but over the years Web mining techniques have developed a full fledged support to provide personalized experience for user. Lot of work is done to improvise Web Personalization.

- Denis Parra, Peter Brusilovsky (2015) [4] investigated the role of user's ability to control a personalized systems by implementing and analyzing a new interactive recommender interface, SetFusion. They checked whether allowing the user to control the process of integrating multiple algorithms resulted in increased engagement and a better user experience. They gave an interactive visualization using Venn diagram combining with sliders which provide an efficient visual model for information filtering. Secondly, authors provide a three-dimension evaluation of the user experience i.e. objective metrics, subjective user perception, and behavioral measures.
- Haoyuan Feng, Jin Tian, Harry Jiannan Wang, Minqiang Li (2015) [5] suggest that capturing and understanding user interests are a main part of social media analytics. Users of social media sites generally belong to multiple interest communities, and their interests are constantly changing over time
- Ahmad Hawalah, Maria Fasli (2015) [6] suggest that web personalization systems are used to improve the user experience by providing tailor-made services based on the user's interests and preferences which are usually stored in user profiles and for such systems to remain effective, the profiles need to be able to adapt and reflect the users' changing behavior. In this proposal, authors introduce a set of techniques designed to capture and track user interests and retain dynamic user profiles within a personalization system.

III. MODEL FOR PERSONALIZATION

Predicting user navigation is a quite difficult in the situation when users are not interested to reveal their personal information on which recommendation can possible .For our new proposal of intelligent system, we consider a concept hierarchy based fuzzy logic method to obtain transaction clusters. Transaction clusters are taken as input and maintained in a set C = {C1, C2, …, Ck} of clusters, where each Ci is a subset of transaction say, T, which is a set of user transactions. Each cluster corresponds to a collection of users with same access patterns.
Let A = {T1, T2, T3,……, Tm} be the set of user transactions and
U ={URL1, URL2, URL3,…., URLn} be the set of unique URLs in all the transactions here Ti □ U for I =1 to m.For all Ti,Tj □ A.

Let C1, C2, …., Cm be the transaction clusters obtained from user transaction A with the help of fuzzy logic based similarity relation.

STEP-1
For the given input i.e. user transactions, accessed URLs and clusters, calculate value of URLi in Cluster Ck. Thakur Bhawesh kumar, Abbas S.Q & Beg Rizwan (2014) [7]. The value of URLi in Cluster Ck is defined as.
The Value of URLi in cluster Ck is the proportion of occurrence cardinality of given URL among transactions in the cluster to the transactions cardinality in the cluster [7].

$$Val(URLi, Ck) = \frac{|\{T|URLi \in Child(T), T \in Ck\}|}{|\sum T \{Child(T)| T \in Ck\}|} \quad \text{Eq. (1)}$$

Where Child (T) is the set of children of transaction T.
The measure of value of an URL in a cluster gives the information on the weight of that URL in that cluster.
Let a new transaction is started, say Tnew. Assign Tnew itself a cluster, that is, Cnew = {Tnew}

ISSN 2348 – 8034
Impact Factor- 5.070

The potential usefulness of Ck with respect to Cnew is a factor defining its interestingness can be estimated by a utility function such as support.

STEP-2
The support of new a cluster Cnew to cluster Ci is defined as

$$\text{Support}(Cnew, Ci) = \frac{|\sum URL(Val(URL,Cnew)-Val(URL,Ci)|}{|\{ Child(T) \mid T \in Ci \cup Cnew \}|} \quad \text{Eq. (2)}$$

Where Child(T) is the set of children of transaction T. Val(URL,Cnew) is the value of URL in the cluster Cnew and val(URL,Ci) is the value of URL in the cluster Ci.
Once a new user starts a session, the objective is to match, at each step, the partial user session with the appropriate clusters and provide recommendations to the user.

STEP-3
The match score between Ci and Cnew is defined as
Match(Cnew,Ci) = 1- Support(Cnew,Ci) Eq. (3)

The preference set of URLs can be made from the highest match score clusters down to lowest match score cluster. We can also impose a minimum threshold on the matching score to reduce the dimensional space.

IV. IMPLEMENTATION

We used the Web Usage access logs from the Web server for experiment. The log data from accesses to the server during a period of 15 days is used for the purpose of recommendation generation.

In order to evaluate the recommendation effectiveness for given model, we measured the performance using three different standard measures, namely, precision, coverage, and the F1 measure. These measures are adaptations of the standard measures, precision and recall, often used in information retrieval.

Traditionally, evaluation of information retrieval system performance is based on the system's ability to retrieve many relevant documents and to not retrieve irrelevant documents The ability to retrieve relevant documents is determined by recall and the ability to avoid retrieving irrelevant documents is determined by precision.

Data Set	Window Size	Precision	Coverage	F1Measure
Data Set 1	2	0.33	1.0	0.4962
Data Set 2	2	0.5714	1.0	0.7261146
Data Set 3	2	0.4705	1.0	0.6394558

V. CONCLUSION

The model for generation of recommendation set for web personalization using clusters of web usage data based on fuzzy logic was evaluated for its effectiveness using standard measures of information retrieval i.e. Precision, Coverage and F1 measure on the data sets

The model is capable of producing recommendation sets with high coverage and high F1 measure with moderate precision

ISSN 2348 – 8034
Impact Factor- 5.070

REFERENCES

1. Arotaritei, Dragos and Mitra, Sushmita, (2004). Web mining: a survey in the fuzzy framework. ELSEVIER Fuzzy Sets and Systems 148, pp 5–19.
2. MULVENNA, M. D., ANAND, S. S., AND BUCHNER, A. G. (2000). Personalization on the net using web mining. Commun. ACM, 43, 8 (August), pp 123–125.
3. Anand, Sarabjot singh and Mobasher, Bomshad, (2005). Intelligent Techniques for Web Personalization. Springer Berlin, Heidelberg, pp 1-36.
4. Denis Parra, Peter Brusilovsky (2015). User-controllable personalization: A case study with Set Fusion. International Journal of Human-Computer Studies, Elsevier, Volume 78, June 2015, Pages 43–67.
5. Haoyuan Feng, Jin Tian, Harry Jiannan Wang, Minqiang Li (2015). Personalized recommendations based on time-weighted overlapping community detection. Journal of Information & Management, Elsevier, Volume 52, Issue 7, November 2015, Pages 789–800.
6. Hawalah, A and Fasli, M., (2015). A hybrid re-ranking algorithm based on ontological user profile., IEEE Computer Science and Electronic Engineering Conference (CEEC), pp 50-55. Thakur Bhawesh kumar,Abbas S.Q, Beg Rizwan., (2014). Web Personalization Using Clustering Of Web Usage Data., IJFCST, Vol.4, No.5, September 2014, pp 69-84.