International Journal of Computer Applications (0975 – 8887)
Volume 105 – No. 13, November 2014

User Profile Mining and Personalization of Web Services

Preeti Khare
Assistant Professor (Computer Engineering Department)
SGSITS Indore (M.P)

Vandan Tewari
Assistant Professor (Computer Technology & Applications)
SGSITS Indore (M.P)

Nirmal Dagdee
Professor Computer Engineering Department)
SDBCE Indore (M.P)

ABSTRACT
Web Services are defined as software systems designed to support interoperable machine-to-machine interaction over a network using standardized XML messages. Since user's expectations and requirements constantly change, it is important to include their preferences in offering of web services. A user profile, used in web service personalization, is a structured construct containing information both directly and indirectly pertaining to a user's preferences, behavior and context. Effective personalization requires services to build and maintain accurate models of a customer's preferences, interests and background through a user profile. Building effective user profiles can benefit from different research contributions in different areas, including security, statistical prediction and mining etc. In this paper we focus on the dynamic evaluation of user profiles for personalization of web services based on service usage log.

Keywords
User Profile, Web Services, Personalization and Service Usage Log

1. INTRODUCTION
Personalization is the process of presenting the right information to the right user at the right moment. User profiles [1] can be defined as a representation of the knowledge about the user's interesting information. The user profile contains different information like basic information which refers to the name, age, address, etc., knowledge of the user which is extracted generally from his web page navigation, interests which are defined through a set of keywords or logical expression, history or feedback which design collected information form user's activity etc.

Behavior of user changes over time. This fact is not taken into account by existing systems which use initial profile provided by the user and provides undesirable services. Currently web service discovery is based on keywords supplied by the user. If a web service crawler engine is used for discovery, multiple services would be discovered as a result. But a large number of discovered services may be irrelevant to the user's requirements. The user therefore has to invoke each and every service until the desired service is fetched. This results in lots of wastage of time and effort. This problem can be solved if additional information about a user is stored and then the search results are personalized, which can be effectively done using user profiles. Typically, there are two types of user profiles: Static (or explicit, obvious or extended) user profiles [2] – Such profiles are built on information provided by the user in the registration or subscription forms of a website. Information such as demographic (location, country), personal (name, age, education) and behavioral (such as preferences) may be used for creating individual or group user profiles. The other type of profiles are Dynamic[3] (or implicit or less obvious) user profiles – Such profiles are inferred by the click stream analysis of a user's behavior based on their activities extracted from web server logs or by the agent that monitors them. This paper proposes a novel method which uses dynamic user profile for personalized web service offerings.

.

2. MOTIVATION
In current scenario of web service usage, the users prefer to receive personalized services. Personalization [4] is the way to offer services directly to the immediate requirements of the user. Because of today's wide variety of services offered to perform a particular task, it is essential that users are supported in the eventual selection of appropriate services. For this reason web services personalization has become important in service discovery. Existing methods for building user-profile have many limitations. The major problem with these approaches is that evolution of user-profile is ignored which disables the dynamic evolution of user profile. Hence the dynamic evolution of user profile can be used to improve the web service selection.

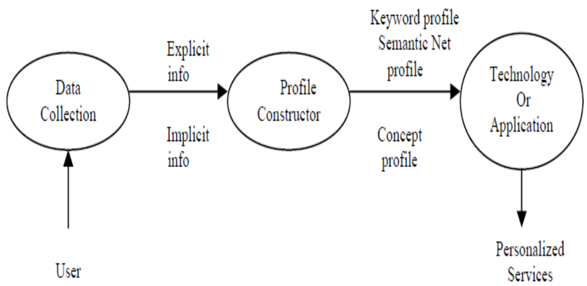

Fig 1: Overview of User-Profile-based Personalization

As shown in Fig.1, the User profiling process generally consists of three main phases. First, an information collection process is used to gather raw information about the user. The second phase focuses on user profile construction from the user data. The final phase, in which a technology or application exploits the information in the user profile in order to provide personalized services.

3. PROPOSED METHODOLOGY
In this paper a sample data set and service usage data has been taken and Modified Fuzzy c-means clustering technique[5] has been applied to form fuzzy clusters of users and on the basis of these fuzzy clusters, results are personalized.

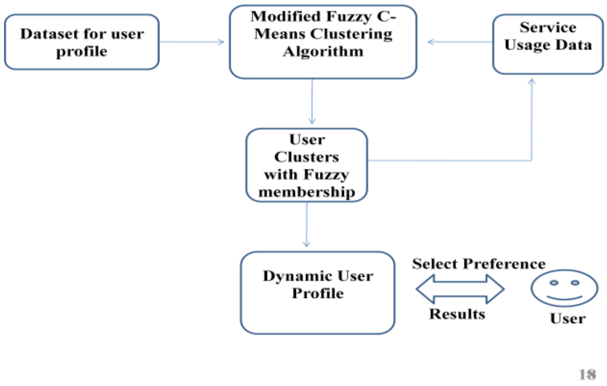

Fig 2: Architecture for Proposed System

3.1 Modified Fuzzy C-Means Clustering Algorithm

Fuzzy c-means (FCM) [7][8] is a data clustering technique in which a data set is grouped into n clusters with every data point in the dataset belonging to every cluster will have a high degree of belonging or membership to that cluster and another data point that lies far away from the center of a cluster will have a low degree of belonging or membership to that cluster. It is based on minimization of the following objective function:

$$J_q = \sum_{i=1}^{N}\sum_{j=1}^{C}\left(u_j(x_i)\right)^q \left\|x_i - c_j\right\|^2, 1 \leq q < \infty.....1$$

Here q is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.

The FCM algorithm and its derivatives have the iterative nature and their calculation often involves a huge number of membership matrices and candidate cluster centers matrices. To solve these problems, an efficient algorithm for improving the FCM called the Modified FCM [5] algorithm has been used. This algorithm works in two phases. In Phase I, the dataset is partitioned into some small block cells and the original dataset is reduced into a simplified dataset with unit blocks. All patterns in a unit block are replaced by the centroid of these patterns. Then, the large number of patterns in the original dataset is drastically reduced to a small number of unit blocks' centroids, i.e., the simplified dataset. The actual cluster center of this simplified dataset is then formed by the FCM algorithm. In Phase II, the cluster centers initialized using the final cluster centers from Phase I. It has been observed that the execution performance of the MFCM is better than that of the FCM and its derivatives.

To partition the dataset into unit blocks the k-d tree [5] method is used. The splitting priority depends on the scattered degree of data values for each dimension. If one dimension has a higher scattered degree, it has a higher priority to be split. The scattered degree is defined as the distribution range and the standard deviation of the feature. The scattered degree is calculated as:

$$R_i = \frac{(X_{max} - X_{min})}{\sigma_i} \quad , i=1...f$$

Here

R_i: the scattered degree for the i^{th} feature (dimension)

X_{max} : the maximum value of pattern in the i^{th} feature

X_{min} : the minimum value of pattern in the i^{th} feature

σ_i : the standard deviation of all patterns in the i^{th} feature

f : the number of features (dimensions)

After splitting the dataset into unit' blocks, we calculate the centroid C for each unit block that contains some sample patterns. The centroid C represents all sample patterns in this unit block. These centroids C are then used to denote the original dataset. The centroid in the i^{th} unit block is calculated as:

$$C = \frac{LSUB_i}{WUB_i}$$

Here

$LSUB_i$: the linear sum of all points in a unit block.

WUB_i : the number of patterns in a unit block.

Fuzzy C-means Clustering Algorithm

Input:

1) Training dataset X ={(x_{11},y_{11}),(x_{12},y_{12})…..(x_{mn},y_{mn})}
2) Cluster center C= {(p_1,z_1),(p_2,z_2)…(p_n,y_n)} where (p_1,z_1)….(p_n,z_n) some initialized points from training dataset.

Output:

1) Cluster membership degree(U)
2) Stabilized cluster center (C1)

BEGIN

Set iteration=1, fv=0, q=2, t=0.0005

/*we have initialize the value of iteration count(i), the objective function(fv), the fuzzification parameter(m) and the termination criteria(t)*/

Step 1: For each training sample d=1 to v and for each cluster center e=1 to w, Calculate the Euclidean distance between each point to the cluster center.

$$\mathrm{Dist}(d,e) = \sqrt{(d_1 - e_1)^2 + (d_2 - e_2)^2 + \cdots + (d_n - e_n)^2}$$

Then, go to step 2

Step 2: For each training sample d=1 to v and for each cluster center e=1 to w, Calculate the cluster membership

$$U(e.d) = \frac{\left(\frac{1}{d_{de}}\right)^{\frac{1}{q-1}}}{\sum_{k=1}^{c}\left(\frac{1}{d_{ki}}\right)^{\frac{1}{q-1}}}$$

Then, go to step 3

Step 3: For each training sample d=1 to v and for each cluster center e=1 to w, Check the constraint the sum of membership degree for each point in sample space must be equal to1

$$\text{If } \sum_{d=1}^{c} u_e(x_d) = 1 \text{ then,}$$

Go to step 4

Step 4: For each training sample d=1 to v and for each cluster center e=1 to w, find the new centroid corresponding to every cluster

$$C1(e,d) = \frac{\sum_{d=1}^{N}[u_e(x_d)]^2 x_i}{\sum_{d=1}^{N}[u_e(x_d)]^2}$$

Then, go to step 5

Step 5: Evaluate the value of the objective function J_q.

$$J_q = \sum_{i=1}^{N}\sum_{d=1}^{C}\left(u_d(x_i)\right) q||x_i - c1_d||^2, 1 \le q < \infty$$

Then, go to step 6.

Step 6: Compute, the difference between the initialized objective function with the new computed objective function

$$\text{objfun} = ||J_q - \text{fv}||$$

Then, go to step 7.

Step 7: Compute, the difference between the initialized cluster center and the new updated centroid.

Center_diff= $|| C(e,d) - C1(e,d)||$

Go to step 8

Step 8: if objfun $< t$ and center_diff $= 0$

Then, stop and Return cluster center (C1) and cluster membership matrix (U)

else Go to step 9

Step 9: For each cluster center l=1 to w, update the old cluster center with the new computed cluster center.

$$C(l,j) = C1(e,d)$$

Then, go to step 10

Step 10: Update, the value of objective function

$$\text{fv} = J_q$$

Go to step 11

Step 11: Update, iteration=iteration+1

Go to step 12

Step 12: Repeat step 1-11 by iteratively updating the Cluster center(C) and the objective function.

4. EXPERIMENTAL EVALUATIONS

The proposed system has been tested on live data of an institute library and by collecting logs of different service provides on a single platform. Dataset contains six attributes such as user id, branch, preferences, age and access frequency, service name.

In the proposed work, the profiles considered are semi-dynamic in nature since a registration process is used to generate explicit user profile. However, these profiles have been suitably augmented using implicit information regarding service usage behavior of a user.

Further, the clustering algorithm has been modified so as to decrease the number of iteration and time complexity of the proposed approach as can be seen from table 1.

Table 1: Result for FCM algorithm

Algorithm	No. of iterations	Total Time	Objective Function
Simple FCM	48	0.98622 sec	0.000590
MFCM	22	0.224644 sec	0.000335

It is visible from these results that the MFCM algorithm has better performance in term of execution time than the FCM algorithm. It reduces the computation cost and improves the performance by finding a good set of initial cluster centers.

Also the service offering using the proposed technique has more precision than simple search due to use of user clusters information. Table number 2 shows the improved precision in the proposed approach

Table 2: Result for Precision

No. of queries	Retrieved Queries	Relevant Queries
10	6	3
20	13	7
30	20	12
40	28	18
50	36	24
60	44	31
70	52	38
80	61	46
90	70	56
100	8	66

As the number of queries increase the respective precision of web services result increase with increases in usage of services. The proposed system has been tested by executing 100 queries by different group users. The access result has been provided on the basis of requirement. When these 100 access requests are made it is observed that 79 results were retrieved correctly and 66 were relevant result in them.

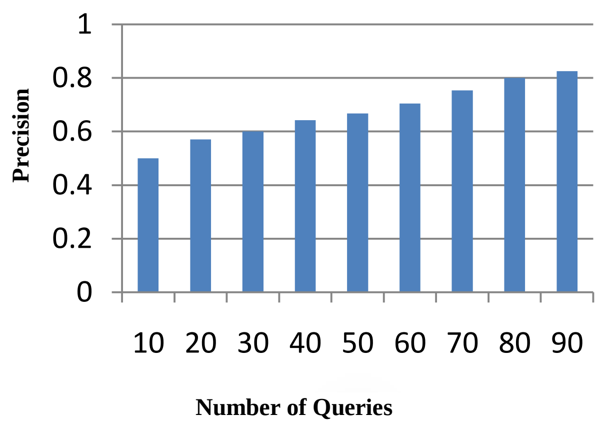

Fig 3: Graph to show no. of queries verses Precision for all services

5. CONCLUSIONAND DISCUSSION

In this paper dynamic user profile mining system has been constructed to personalize web service discovery. The proposed system is based on dynamic user profile mining for personalization. The following inferences have been drawn based on the observations taken during the development of the proposed system:

- By using the dynamic profile generation, the web service search result improves in precision.
- Using past service usage log benefits the user and improves service discovery.
- The FCM algorithm and its derivatives have iterative nature and their calculations often involve a huge number of membership matrices and candidate cluster center matrices. It is a computationally intensive method. However the modified FCM algorithm performs better than the FCM algorithm as it reduces the number of iterations and hence improves the performance by finding a good set of initial cluster centers.

In future this work can be augmented by using other algorithms for clustering the user profiles and their efficiency may be observed. Other aspects of personalization can be focused upon such as a multilevel recommender system can be used which emphasize on different levels of similarity between profiles. The ever growing need for personalization in discovery of web services can also focus on non functional preferences of the user in addition.

6. REFERENCES

[1] K.Sugiyama,K.Hatano Et Al,Adaptive “Web Search Based On The User Profile Constructed Without Any Effort From Users”.

[2] M. Bonett, “Personalization of Web Services: Opportunities and Challenges”, ARIADNE, 28, June 2001.

[3] Baraglia R., Silvestri F.:Dynamic Personalization of Web Sites Without User Intervention. Comm. ACM, Vol. 50, 2 (Feb., 2007): 63-67

[4] Magdalini Eirinaki and Michalis Vazirgiannis “Web Mining for Web Personalization”, Athens University of Economics and Business, 2003.

[5] Ming-Chuan Hung and Don-Lin Yang, “An Efficient Fuzzy C-Means Clustering Algorithm” Department of Information Engineering, Feng Chain University Taichung Taiwan.

[6] Keno Buss, “Literature Review on Preprocessing for Text Mining,” STRL, De Montfort University.

[7] J.C.Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York 1981

[8] S. Akthar, Sk. Md. Rafiet, Improving the software architecture therough fuzzy clustering technique, Indian Journal of Computer Science and Engineering1(1), 54-57.