Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PROVIDER RECOMMENDATION IN E-SERVICE PORTALS USING GROUP PROFILING

HEBAH ELGIBREEN[1], ABDULRAHMAN MIRZA[2]

[1]IT Department, College of Computer and Information Sciences, King Saud University

[2]IS Department, College of Computer and Information Sciences, King Saud University

E-mail: [1] hjibreen@ksu.edu.sa, [2] amirza@ksu.edu.sa

ABSTRACT

Nowadays, most of e-service portals focus on individual users rather than considering shared characteristics that could be useful when searching a portal. Service providers searching activities involve different types of common user behaviors and preferences that should be considered. Thus, group profiling can be used to group customers with similar interests together and recommend service providers based on their common characteristics. Accordingly, this paper proposes a new approach that recommends service providers, based on group profiling and ranking, to be used by e-service portals. Such an approach is unsupervised and is carried-out in two steps. First, group profiles are generated. Second, whenever a user tries to search for a service provider, a different result may be returned based on the data collected about that specific user. Thus, customers would most likely find the most suitable service provider more accurately and in less time. This approach focuses on group profiling and group users based on the Ant Colony Clustering (ACC) method, which is a new type of clustering that is based on an Artificial Intelligence algorithm called Ant Colony Optimization. In addition, this approach has been tested and three measures have been recorded including speed, aggregation precision, and result accuracy. Such test has shown that this approach is promising and produces high quality results.

Key words: *Group Profiling, Search Recommender, E-Service Portal, Evolutionary Algorithm*

1. INTRODUCTION

Due to the growth of internet usage, e-services have been increasing in popularity. E-service has been defined as *"an entity available via the Internet that completes tasks, solves problems, or conducts transactions"* [1]. It has been stated, in [2], that the number of customers communicating with organizations and conducting online transactions through e-services is increasing dramatically every year. As a result, portals have been used to facilitate customers' needs through e-services. Portals have been defined as one single point of access or gateway for different services through the web [3]. Consequently, an e-service portal is a gateway for services that are available online. Such gateway hosts different service providers that offer different or similar services.

Portals offer different service categories for different providers [3]. For business success, it has been found that portals must respect customer demands and personalize the offered services [4]. Thus, a major problem that customers face when accessing an e-service portal is selecting the best provider for their needs, especially if more than one provider offers the same service. A customer must search for all providers, compare their prices, check older customers' reviews, etc. Such process consumes a lot of time and energy in addition to the possibility of not choosing the most suitable provider. To solve such problem, a method called profiling can be used. Profiles have been defined as a representation of user behavior that is organized and managed in a way that is suitable for the required domain; such that, it is a knowledge that is formed as a result of turning data into information, to distinguish between relevant and irrelevant behavior [5].

Profiling has been divided into different types, depending on the output requirement. However, since the motive of using profiles in e-service portal is to find common features between customers and categorizing them, then a group profiling method should be used. A group profile *"identifies and represents a group (community or category), of which it describes a set of attributes"* [5].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Besides that, recommendation is considered as a typical personalization technique, which considers different users preferences and needs [6]. A personalized recommendation is a key tool in website services. It has been stated in [7] that personal services are significant to the success of the current business field; thus, recommendations improve the growth of e-commerce marketing activities. In addition, since e-service portal recommendations will present different providers depending on different tastes, it will reduce searching for interesting providers and also enhance e-commerce by increasing cross-selling, converting browsers into buyers, and increasing customers loyalty [7].

Using profiling in recommendations will grant a general overview of the massive data available. It will facilitate information retrieval and simplify the searching process, and, thus, personalization. Moreover, using profiles will ensure achieving the required goal efficiently and easily [8]. To further motivate the use of group profiling, group profile needs less information than individual profiles. In addition, group profile data is usually anonymous and, thus, its individual's information is protected more than individual profiles.

Consequently, the aim of this paper is to use group profiling in an e-service portal to group different customers based on their similar characteristics. Based on the usage observations of each group and its evaluation, the provider of each service will be ranked; such that when a new customer who shares the same characteristic of a certain group search for a service provider then, based on his/her group ranking, a recommended result will be returned. Thus, customers would most likely find the appropriate provider more accurately and in a shorter time period.

The proposed approach acquires and refines different kinds of data explicitly and implicitly. Then, it clusters the users based on the Ant Colony Clustering method in order to aggregate their characteristic later on. Finally, group profiles will be generated in XML files to be used later in the matching process when a user sends a query. The details of such approach are discussed later in this paper, and its performance is tested and evaluated. It must be noted that in this paper the focus will be more on group profiling because recommendation and filtering have a rich background and different improvements have been made in these areas while, on the other hand, group profiling is still emerging and not a lot of works have been done in that area. Hence, it can be said that the proposed approach will contribute in reducing the time of profile matching, improve the accuracy of searching e-services, and increase security.

Finally, it must be noted that this paper is an extended version of a short paper that have been published as a work in progress in [9]. However, sufficient details of the application, method, analyses and results achieved will be indicated and explained here.

This paper is organized as follows. First, general background will be presented and the main aspects of the topic will be discussed. Then, some related work will be presented and discussed. After that, the details of the proposed approach will be discussed and explained. The evaluation of the approach will also be presented and clarified. Finally, this paper will be concluded and future work will be discussed.

2. BACKGROUND

In this section, the main aspects that are needed to understand this paper are defined and discussed. First, e-services, portals, and personalization are defined and explained, then the group recommendation is identified, at the end, user profiling and evolutionary algorithms are presented and further discussed.

2.1 E-Services

E-service can be defined as "*interactive services that are delivered on the Internet using advanced telecommunications, information, and multimedia technologies*" [10]. In order to accurately deliver such services, combination between the Web, legacy systems, and end-users system is necessary [11]. E-services increase the efficiency of the online economy and guide to a new revenue flow. It can be distinguished by a number of characteristics, as follows. It is accessed through the Web and has a particular URL. It is usually composed from different services, and each e-service might depend on other e-services. Furthermore, they are implemented differently and, thus, its management is difficult. Finally, protocols agreement is necessary for communication between different e-services [1].

2.2 Portal

After the mid-1990's, in which portals spread on the Web was detected, portals' functionality has developed extensively. At the beginning portals were more of search engines, but they expanded afterwards and quickly matured to improve their functionality. Advanced searching capabilities were

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

added, contents were enriched, and users were offered more control [12]. A portal offers the possibility to access different types of services and providers [3]. Thus, it can be defined as a gateway or central access to different services available through the Web. It integrates content from different sources and provides personalization features to the users [13].

2.3 Personalization

E-service personalization is *"the process of getting web users' information online, and using the information to tailor web pages to individual users' preferences and deliver services to the user's needs"* [14]. Personalization is capable of recognizing users when they access their account to adopt the presentation of specific content [6]. Recommendation is considered as a typical personalization technique. Therefore, recommendation seeks to predict interesting items on behalf of the user, automatically, according to his/her preferences. Hence, an e-service portal recommendation will present different providers depending on different tastes.

2.4 Group Recommendation

Group recommendation acquires individual preferences of a group member and recommends a certain solution based on the gathered information of all individuals. As a result, aggregation of individuals' information is needed in order to combine their preferences and suggest the suitable solution for all the group members. It has been stated in [15] that regardless of the technique used to gather and represent group individuals, all approaches would be using one of three main schemas of aggregation or filtering. It either aggregate individual ratings, merge their recommendations, or to construct a group preference model. Each schema has it pros and cons, however, depending on the needs one of these schemas can be chosen; more details are further discussed in [15]. The use of group recommendation will personalize e-service portals depending on each user needs and preferences. In order to achieve such a process user profiling is necessary.

2.5 User Profiling

The main motivation of building a user profile is that users exploit the available services differently. Discovering users' differences is essential to provide the required personalized service. Consequently, profiles can simply be defined as a representation of user behavior that is organized and managed in a way that is suitable for the required domain. In general, to profile a user, data must be collected and stored as profile content. Thus, the process of user profile construction has been divided and identified in [16]. First, data is collected and gathered from the resource available. Then, collected data are stored and aggregated in the database. Afterwards, stored data will be examined in order to extract the result and then interpret it. At the end, extracted knowledge will be applied on new activities in order to improve the studied problem. In addition, in order to build and use profiles, different techniques have been developed and used, such as Evolutionary Algorithms [17]. Furthermore, some authors have tried to combine more than one technique to build user profiles more efficiently.

2.6 Evolutionary Algorithms

Evolutionary Algorithms (EA) [17] are Artificial Intelligent algorithms that have been inspired by nature. Such algorithms are used to solve any kind of problem and identify candidate solutions as fast as possible. It is useful when the best solution is unknown and the data repository is very large. Different techniques have been developed under the umbrella of EA. Each technique has been tailored to different domains depending on the problem and solution required. In this paper, the focus will be on the Ant Colony Clustering (ACC) method [18] for problem solving.

The ACC method is a new type of clustering that is based on an Evolutionary Algorithm called Ant Colony Optimization, which is built based on nature. In specific, it is developed based on how ants look for food by spreading their pheromone to find the food path [19]. Figure 1 illustrates how ants search for food. They start updating their local pheromone [Figure 1.b] and explore all possible roads to finally emphasize the shortest one. As a result, the global pheromone [Figure 1.c] will be emphasized, while the other pheromone is faded, by communicating the local pheromone, and hence, the best solution will be chosen.

Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

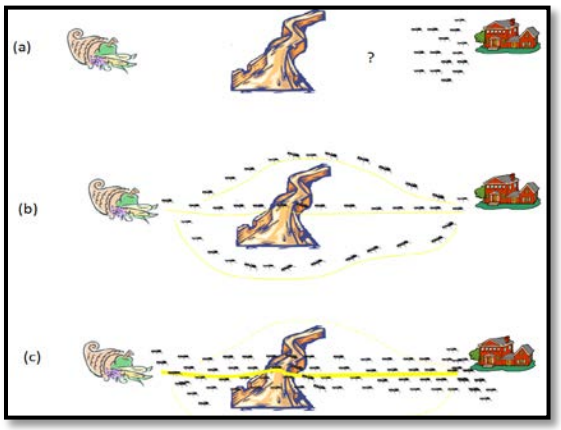

Figure 1: Ant Colony Optimization

The ACC is developed to optimize the process of clustering and improving its time. It is mostly used in web usage mining [14] and basically divided into three steps [20]. First, it updates the local pheromone, i.e. update each ant solution with the suitable cluster. Second, it applies the state transition rule, i.e. decide if a certain ant should be dropped in a cluster or out of it. Finally, it updates the global pheromone, i.e. update the road map which is the final solution.

3. RELATED WORK

Taha and Elmasri [21] proposed a recommender system called SPGProfile. They used ontology-based social networks to group users based on their ethnicity, demography, culture, age, religion, and other characteristics. Their system filtered and ranked a user searching result based on the preferences of the social group. In addition, Haewoon et al.[22] proposed a supervised method that groups users based on their Web tag, which represent users' interests. The objective of their method was to group similar users across multiple services, specifically on some of the popular web services.

On the other hand, Gomes and Canuto [23] presented a system, called Caracará, which uses user profiles and dynamic grouping processes to facilitate Web searching. Their system monitors users' actions (accessed URLs) to dynamically suggest the result list based on similar users who belong to the same group. Furthermore, in [24] ontology concepts, professions list, and targets' URL were used to build group profiles. Their profiling is used to customize semantic browsing in a medical research portal.

Moreover, in [25] a recommender system was proposed to improve the efficiency of recommendations using time context and group preferences. However, Senot et al. [26] tried to determine what factors influence the choice of grouping strategy. They considered different types of content, including individual characteristic, context, and group interactions tracing to evaluate their approach using a real large-scale dataset of TV viewings. Furthermore, in [27] different strategies have been proposed to group ontology-based user profiles into one semantic profile. Their profile contained individual preferences and were evaluated empirically and theoretically in a knowledge-based multimedia retrieval system.

Last but not least, work was conducted in [28] to identify the relationship between groups of people in a scientific social network. Their clustering technique identified and grouped research communities in the Brazilian scientific social network. Furthermore, in [29] a clustering technique was conducted to group similar consumer preferences and business-related information in a telecom service. Such profiling enabled telecom service personalization and, thus, increased the capacity of their services.

From all the work discussed above, it can be noticed that most of the available research in recommender systems concentrates mostly on the filtering process. Specifically, in e-services, behavior and characteristics of users were not considered when recommending similar searching results. They only considered what key words the users used when searching, while characteristics, like age, and behavior, like spending habits, have been neglected. Moreover, even if profiling is used they usually profile an individual and then, during the filtering stage, grouping is applied. Such process wastes a lot of time because grouping will be conducted every time the user sends a query. It also causes some space and security problems since all individual profiles must be maintained. As a result, the next section will present a new approach that solves such problems by using group profiling techniques to reduce time of profile matching, improve accuracy of searching e-services, and increasing security.

4. E-SERVICE PORTAL RECOMMENDER

After understanding how recommendation and group profiling have been applied in the field of e-service portals, this section describes the new approach which uses group profiling in e-service portals to personalize and recommend the appropriate service providers for the customers. Specifically, an unsupervised approach is applied,

Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where previous knowledge or examples are not given in advance so that the recommender can learn from; hence, the recommender learn from scratch without samples that give him a hint of how to recommend. This guarantee that any behavior can be considered and to simplify the process of profiling since there is no need for samples to be collected or expert users to be hired. The proposed approach is mainly divided into two steps. First group profiles must be generated, then, whenever a user tries to search for a provider, a result will be recommended depending on the data collected.

4.1 Group Profiling

In this step, customers will be profiled in groups, and providers will be ranked differently in each group. For each service, the step of group profiling is illustrated in Figure 2 and the details of each step are described next.

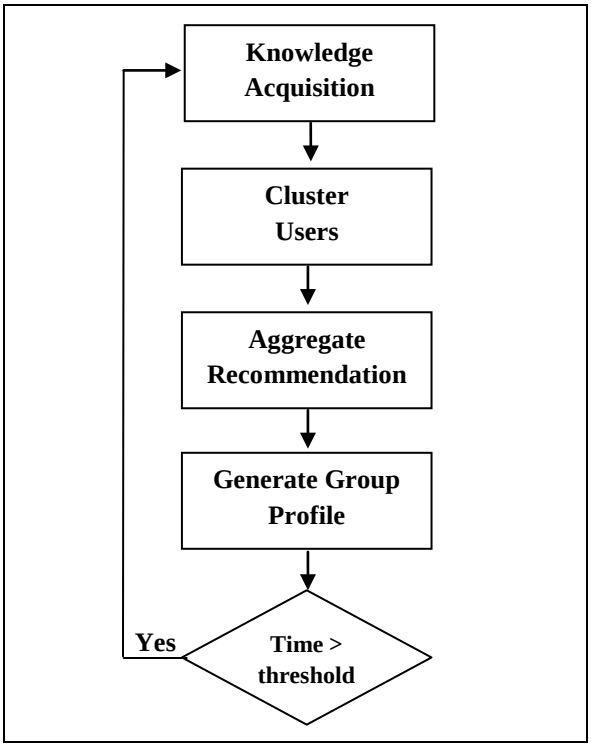

Figure 2: Group profiling process

A. Knowledge Acquisition

The first step of the proposed approach is to acquire the knowledge required to profile users and build the groups' profile. Such step is further divided into two main phases: collection and refinement. The data must be first collected from different sources explicitly and implicitly. It is a mixture of four types, including: behavior, demographic, characteristics, and rating data. Behavior data contains the spending and the usage behavior of the users, and it is implicitly extracted from the financial transactions that are carried-out through the portal. On the other hand, demographic or context data contains the country of the customer and the time of purchase. Moreover, characteristics' data includes the date of birth and the education level of the customer. Finally, the ranking data includes the evaluation that a customer has given for developers. All these data will be collected directly/indirectly from the user and his transaction.

After that, in order to use and extract the required knowledge and to build user profiles, a refinement process is necessary. Such step will further generalize the collected attribute and simplify its values based on a data mining characterization approach, in specific Attribute-Oriented Induction approach [30]. As a consequence of this step the time attribute will be generalized to (Morning, Afternoon, Night), age to (Kid, Adult, Elderly), spending and service usage to (Low, Medium, High). Thus, each user will have a table that contains the knowledge acquired in addition to the service and provider identification number. An example of the acquired table, of one user, is illustrated in Table 1.

Table 1: Acquired knowledge of a user

Service	Computer maintenance	House keeping
Provider ID	2	10
Country	Saudi Arabia	Saudi Arabia
Age	Elderly	Elderly
Education	Master	Master
Time	Morning	Night
Spending	High	Low
Usage	Low	High
Rank (1-10)	4	1

B. User Clustering

After acquiring the knowledge needed for profiling, the second step is to cluster each user and conduct the group profile. The proposed approach clusters the users inclusively; such that clusters (or groups) might have shared user, i.e., the user might belong to more than one group. This is because a user could evaluate the same provider differently at different times. The clustering is based on Ant Colony Clustering (ACC) method [18]; which is basically divided into three steps: local pheromone update, state transition rule, and global pheromone update [31]. In our knowledge, no one yet has used ACC with hamming distance, which count how many symbol between two string are different, to build group profile, especially with hybrid content.

Journal of Theoretical and Applied Information Technology
15 August 2012. Vol. 42 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Each user will represent an ant, and the clusters will represent the road that the ant should take to get the best possible solution, i.e. accurate clustering. The solution is represented by a vector with N elements and W values; such that N represents the number of customers' evaluation and W represents the number of clusters. For example, if the number of customers evaluating were four and the number of cluster given are three then a solution could be (2, 1, 3, 1) which means that the second and fourth customers' evaluation belongs to the same cluster while the first belong to cluster # 2 and the third belong to cluster # 3.

After initializing all the attributes, the proposed approach works as follows. First each customer evaluation will be assigned randomly to a cluster, i.e., ants are positioned randomly on the road. Then, while the stop condition is not satisfied, each customer is matched with all other customers, i.e., each ant searches its neighborhood to find the most similar ant. The similarity measure used in the proposed approach is based on hamming distance but, instead of matching character string, the attributes are matched. The similarity measure between customer Xi and Xj is calculated using Equation (1), where U is the total number of characteristics, and AttMatch is the match between attributes of two customers and calculated using Equation (2).

$$f_{(i,j)} = \frac{1}{U} \sum_{h=1}^{U} \boldsymbol{AttMatch}(X_{ih}, X_{jh}) \quad (1)$$

$$\boldsymbol{AttMatch}(X_{ih}, X_{jh}) = \begin{cases} 1 & \text{if } X_{ih} = X_{jh} \\ 0 & \text{if } X_{ih} \neq X_{jh} \end{cases} \quad (2)$$

After calculating the similarity measure it will be possible to apply the transition rule in order to update the local pheromone. Thus, the probability of picking up (P_P) a customer cluster or dropping it down (P_d) must be calculated. Such probability is calculated using Equation (3) and (4) where f is the similarity measure, K_p and K_d are threshold constant defined at the beginning.

$$P_P = \left(\frac{K_p}{K_p + f_{(i,j)}}\right)^2 \quad (3)$$

$$P_d = \left(\frac{f_{(i,j)}}{K_d + f_{(i,j)}}\right)^2 \quad (4)$$

Usually, in the original ACC algorithm, if the current ant does not carry an object (cluster) while its neighborhood has an object, then it will take the same object as the other ants based on the pickup probability, i.e., join their cluster. On the other hand, if the current ant found to be carrying an object but its neighborhood carries a different object then it will drop its object based on a drop down probability, i.e., it will not belong to any cluster. However, it has been found that such manner in the proposed approach affects the clustering accuracy, especially if the global solution was initialized with no cluster. Thus, a slightly different transition rule has been exploited and used to improve the accuracy, as follows. If the current customer has a similar clustered neighborhood then the neighborhood cluster will be picked up, based on the pickup probability P_p, and the current customer will join the same cluster. On the other hand, if the current customer has a similar cluster to non neighborhood customer this means that the clustering is inaccurate in either current or neighborhood customer; therefore, the target customer must drop down its cluster, i.e., will not belong to a cluster, based on drop down probability P_d.

At the end, after comparing all customers together and update their local pheromone, the global pheromone will be updated. Then, the process will be repeated in order to further improve the solution and assign the customers to the most appropriate cluster or group. It must be noted that only one ant can update the global pheromone in each round. In addition some ants might not find any similar neighborhood, thus, in such a state it has been decided to leave such a customer in a separate cluster to make sure that his characteristic does not disappear, especially since all customers of a cluster will be aggregated in one profile. Doing such a step will make sure that unique customers profile is preserved, and their characteristics are not omitted, to improve the clustering in case a similar customer has joined the portal. The pseudo code of the proposed clustering algorithm is illustrated in Figure 3.

ACC Algorithm {
Input: Acquired Knowledge (X), K_p, K_d, Number of Cluster (NC), Stop Condition (α).
Output: Solution cluster (C).

Initialize global pheromone C
// spread the ants
For each customer do
    Assign cluster randomly based on NC;
End for
a = 0;

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

```
// start the learning process
While ( a <= α ) do
    i = j = 0;
    //apply transition rule and update local pheromone
    For each customer do
        i++;
        for all other customers do
            j++;
            Calculate f(i,j);
            //if more than 75% of the characteristic matches,
              then similar
            if (f(i,j) >= (0.75) then
                Calculate Pp;
                Update Xi local pheromone with Xj cluster (if it
                has one) based on Pp
            //if not similar but same cluster then drop
            Else if (Xi cluster = Xj cluster) then
                Calculate Pd;
                Drop down Xi cluster based on Pd
        End for
    End for

//update stop counter
a++;
//update global solution
Update global pheromone C;
End while
Return C;
}
```

Figure 3: ACC algorithm

C. Aggregate Recommendations

Even though collaborative filtering is mostly used in the recommendation process, it suffers from the problem of cold start [18], where no recommendation is found at the beginning of using the website, and also it reduces the security and increases the time needed to match all customers. As a result, in addition to group profiling, aggregation is used instead of collaborative filtering. This will guarantee customer privacy, reduce matching time, and avoid cold start. Thus, the global pheromone resulting from the clustering process will be considered as the clustering solution; and since the focus of this paper is on group profiling, the ranking process of the providers will be based on ranking aggregation, as follows.

A vector that contains the clusters numbers will be used to aggregate the providers ranking of customers in the same cluster. If the provider is ranked differently from more than one customer then the average of their ranking will be considered. On the other hand, if a provider is not ranked yet then the default ranking will be considered, to avoid the cold start. However, if the default ranking is equal to another evaluated provider, i.e. conflict in the rank, then evaluated provider will come first. Alternatively, if the conflict is between two evaluated providers, then the most recent evaluation will come first then the next one. Finally, as a consequence of this step, a vector of ordered providers will be assigned to each group depending on its customer evaluation. The pseudo code of the aggregation process is illustrated in Figure 4.

Recommendation Aggregation {
Input: customer cluster (C), local ranking list (L), default ranking list (D).
Output: global ranking list (R).

```
//for each cluster ➔ aggregate all its customer profile
For each cluster C do
    i = 0;

    For each provider do
        i++;
        //if provider ID exist in L
        If provider evaluated then
            Update RC (i , L);
        Else
            Update RC (i , D);
        End if
    End for
End for
Return R;
}
```

Figure 4: Recommendation aggregation

D. Generate Group Profile

In order to generalize the proposed approach and increase its scalability the groups' profile is represented and stored in XML files. As a result, such files can be used with any other technique and, thus, it can be applied in any other domain. In addition, it is well known that XML files contain structured data that is represented as tree; thus, searching in this type of structure will be much faster than searching in a relational database or vectors.

For each service, an XML file will be generated to contain the gathered group characteristics; such that the most general customers' characteristics will be chosen. Thus, individual property will be secured and the space

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

will be reduced. In addition, the file will contain the ranking of the providers in order to be used in the recommending process. An example of such file is illustrated in Figure 5.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
- <SupportService name="ComputerMaintenance">
  - <Cluster number="0.1">
      <country>KSA</country>
      <age>elderly</age>
      <education>High school</education>
      <spending>high</spending>
      <usage>low</usage>
      <orderTime>morning</orderTime>
      <provider ID="101">1</provider>
      <provider ID="108">2</provider>
      <provider ID="118">2</provider>
      <provider ID="119">3</provider>
      <provider ID="103">4</provider>
      <provider ID="110">4</provider>
      <provider ID="106">5</provider>
      <provider ID="120">5</provider>
      <provider ID="114">6</provider>
    </Cluster>
  + <Cluster number="0.2">
  + <Cluster number="0.3">
  + <Cluster number="0.4">
```

Figure 5: Group profile file

E. Updating

Even though group profiles are generated, permanent customer interest cannot be guaranteed. It is possible that customers change their mind about a certain provider. Their purchase behavior might change, and their characteristics might be updated. Thus, changing in the long term interest must be considered. In the updating process, a time threshold will be applied to repeat the process of profiling and ranking aggregation. For example, when a customer searches after a couple of weeks of using the same searching criteria, a certain provider might be ranked as first, while he used to be the third.

4.2 E-Service Personalization

When a customer starts his searching process, given the service name, his/her profile must be generated in order to compare it with the groups of the enquired service. If the customer is not registered, then only the place and time of searching can be collected from the Web log files. In addition, an option is offered to the customer to enter his/her characteristics, including the date of birth and education, in order to further improve the matching process. However, if the customer already registered then the missing data can be acquired from the account information.

After building the individual profile, the XML file of the inquired service will be fetched and the similarity measure between the target customer and the groups' profile will be measured, using Equation (1). After that, the most similar group will be chosen and, thus, the providers will be viewed by the customer based on the stored ranking list. The details of this step and its pseudo code are illustrated in Figure 6.

Personalization algorithm {
Input: Group Profiles (G), individual profile (I), service enquired (S).

Output: providers list (P).

$i = F_{old} = F_{new} = Best = 0;$
For each group G
  i++;
  F_{new} = Calculate $f_{(I,Gi)}$;
  If $F_{old} < F_{new}$ then

    // Best group number matches is G_i
    Best = G_i
  End if
End for

P = recommendation list (Best);
Return P;
}

Figure 6: Personalization algorithm

It must be noted that, due to the possibility of preference conflict, all providers will be viewed by the customers. However, the ranking of the providers will differentiate depending on the customer criteria. As a result, the problem of cold start is further reduced and, thus, customers can choose other than what the portal recommends. It can be said that such process is decision support for the customer rather than decision making on behalf of the customer.

5. IMPLEMENTATION

In order to test the performance of the proposed technique an experiment has been conducted. This section discusses the details of the experiments, explain its result, and highlight its main issues.

5.1 Experiment Settings

The proposed approach has been implemented with Java language in JBuilder environment. The experiment was conducted on a PC with Intel®Core™ i7 CPU, 2.67 GHz processes, and 6GB RAM. In addition, a synthetic database has been used with the experiment. Such database contained the customers' information, their ranking of the used providers, providers and service

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

information, in addition to the default ranking of each provider.

During the experiment, all the parameters remain unchanged, except for the stop condition, including: service name, pick up and drop down probability, in addition to the number of clusters. Since e-service portals have hundreds of customers who can give thousands of evaluations for different providers, one service name was chosen through the experiment, which is "computer maintenance". Such service has, in our database, one hundred customers, twenty providers, and one thousand evaluation transactions. In addition, the pickup probability (K_p) has been initialized with "1" in order to increase the resultant P_p and, thus, increase the probability of picking a neighborhood cluster when found. On the other hand, the drop down probability (K_d) has been initialized with "0.5" because it is unknown if it is a good decision to drop the cluster or not and, thereby, a 50% chance has been given to drop down a cluster. Lastly, since a thousand transactions are clustered then the maximum number of group was set to "10".

In the experiment two types of queries have been used. The first one is for a registered customer, i.e., full characteristics are available, while the second is for a new customer, i.e., only time and place are identified. In addition, ten trials have been recorded to measure the performance of the proposed approach; each trial has a different stop condition starting from 1000 to 2000. In each trial different measures have been calculated; in the next section, the used measures are identified and explained.

5.2 Evaluation Measure

In order to evaluate the results of the experiment three types of measures have been calculated. The first measure is the matching speed, which record the time of executing the matching phase of the algorithm, i.e., the personalization step (fetching query result). The second measure, however, records the precision of the profile aggregation. Such that, for each cluster the sum of AttMatch in Equation (2) is calculated between the cluster profile and its customers profile and then divided by the number of customers. At the end, the precision of all clusters is added and divided by the number of clusters in order to scale it from 0 to 1. This measure will assess the percentage of profile aggregation precision, in terms of the characteristic match. The third measure, on the other hand, assesses the result accuracy. Such that the sum of AttMatch in Equation (2) is calculated between the recommended cluster and the target customer to measure how accurate the selected result is. The result of these measures and their details are further discussed in the next section.

5.3 Experiment Results

After conducting the ten trials for the two types of customers, using the parameters specified previously, each measure is calculated, in each trial, and recorded. In Figure 7, the time taken to execute the matching step is illustrated. As it can be noticed the algorithm usually takes a few milliseconds (less than half a second) to retrieve the required result, either when the customer is registered or not. Such result is required because when customers are searching they do not like to wait for few seconds to see the result. In addition, it can be noticed from the figure that the number of iterations has no effect on the matching time; such result is normal because the number of iterations is only used with clustering, which can be done once every period of time on the background without affecting the process of personalization.

Additionally, the aggregation precision between the cluster profile and its customers is illustrated in Figure 8. As it can be noticed, the number of iterations also did not affect the aggregation accuracy. Even though such aggregation depends on the stop condition, but it can be said that a high number of iterations does not necessarily improve the clustering process. Thus, it would be good to choose a stop condition that is equal to the number of evaluations (transaction) to make sure that all local ants have updated the global solution without a needless use of high stop condition.

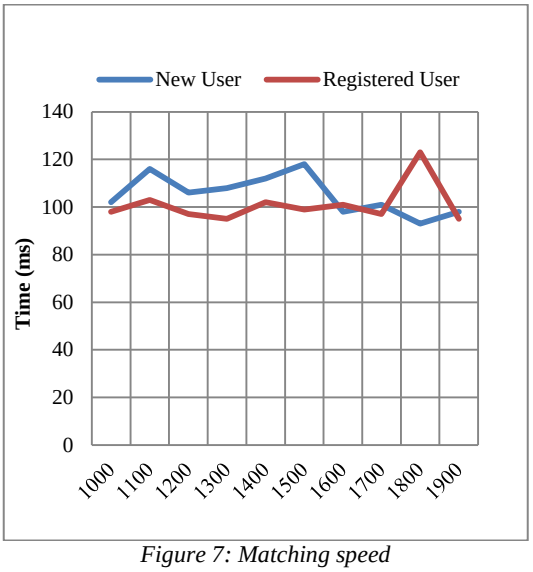

Figure 7: Matching speed

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

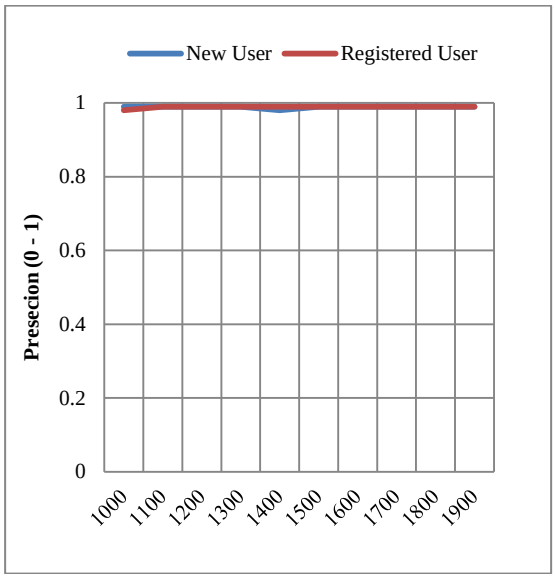

Figure 8: Aggregation precision

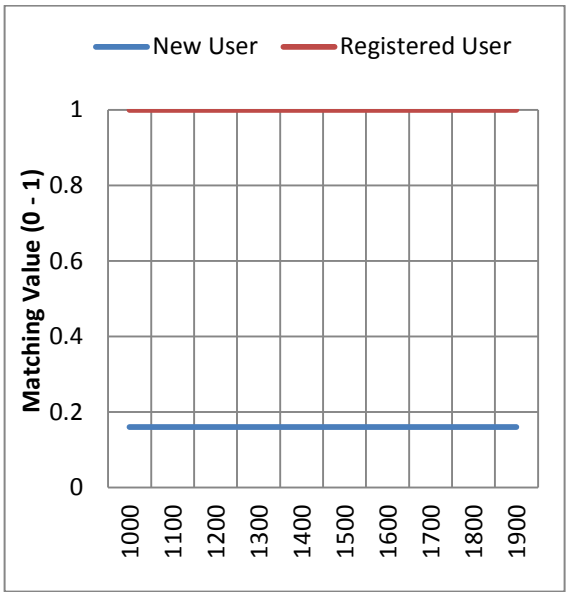

Figure 9: Result accuracy

In addition, it can be noted that the result of the aggregation is almost equal in both cases of registered and new customer. Such result is normal because usually the aggregation is done on the background regardless of the customer requesting the query. Moreover, after recording the resulting clusters, it has been found that such algorithm is very accurate such that the aggregation precision ranges between 98% and 99%, which is a very good percentage.

Finally, the resulting cluster profile has been matched with the target customer profile in each trial, as illustrated in Figure 9. As it can be noticed there is a difference between the result of the registered and the new customer. Such difference is resulting from the lack of information when the customer is new (only two characteristics can be used in the matching process). As a result, it is recommended to get the customer information before the query execution in order to retrieve more accurate results. Moreover, as it can be noticed from Figure 9 that the accuracy of the result when the customer is registered is very high (100%), i.e. a matching profile has been found and extracted.

5.4 Discussion

After understanding and testing the proposed approach, it is time to summarize its contribution and compare it with other works. First of all, such approach considered the long and short term interest; such that it does not always update and re-execute the whole process. Only the matching step is repeated with every query, however the rest is done once and updated every period of time while its result is stored in XML files. Such decision reduced the time tremendously without affecting the accuracy and, also, improved the scalability and accuracy of the recommendation. In addition, the use of XML files has simplified and sped the process of matching because XML files are structured based on tree representation; thus, group profiles are easily searched and information is fetched quickly, much faster than relational databases or vectors.

In addition, when the knowledge is acquired to build a groups' profile, three different types of customer data have been considered: characteristics, behavior, and demographics. This would increase the matching accuracy and improve the result satisfaction. On the other hand, as discussed previously, the available work did not consider the behavior and characteristics of customers, only key words matching of customers log files have been used. Thus, our approach has contributed in this area.

Moreover, even though collaborative filtering is mostly used in the recommendation process, but it suffers from the problem of cold start, reducing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

security, and increasing of time needed to match all customers. In the proposed approach, however, group profiling was used instead, such that aggregation is used instead of collaborative filtering. This will guarantee customer privacy, and reduce matching time. In addition, to further improve the result and avoid the cold start and due to the possibility of preference conflict, all providers are viewed by the customers. However, the ranking of the providers is different depending on the customer criteria. As a result, personalization is considered with the possibility for customers to choose other than what the portal recommends. Thus, such process is decision support for the customer rather than decision making on behalf of the customer. Finally, it must be said that even though the result is mathematically good, but the satisfaction of the customer is not yet assured. Thus, as future work, the proposed approach should be tested on real data with real customers.

6. CONCLUSION

Since most of the existing e-service portals focus on individuals rather than shared circumstances, this paper has proposed a new approach that considers the group behavior in which searching is affected. It is well known that service activities are social and involve different types of individuals. Such individual's behaviors have been considered in order to improve the performance of e-service portals. An e-service portal was defined as a gateway for services that are available online to host different service providers that offer different, or even similar, services. It has been found that portals must respect customer demands for personalization services to improve the market. Thus, a major problem that customers face when accessing an e-service portal is selecting the best provider for their needs.

Consequently, this paper has proposed a new approach which uses group profiling in e-service portals to personalize and recommend the appropriate service providers for the customers. It is an unsupervised approach to guarantee that any behavior can be considered, and dynamic grouping can be applied. Furthermore, the process of profiling is simplified since there is no need for samples to be collected or expert customers to be hired. It is mainly divided into two steps: generating group profiles and recommending of providers depending on the customer sending the query. The first step, however, is further divided into four different steps; which acquire and refine data, cluster the resulting data and aggregate it in group profiles afterwards, then such profile will be generated in XML files to be used later with every searching query. After implementing the proposed approach, it has been found that, for either unregistered or registered customers, the matching speed and aggregation precision result is very good. Such that, the matching process only takes less than half a second while the aggregation precision is, on average, 99% accurate. On the other hand, the result of accuracy test has shown that, in contrast to registered customers, unregistered customers might get inaccurate results due to the lack of information. Thus, it is recommended to urge the customers to enter their information or register in order to improve their searching result. However, in general such approach has shown promising results, and it would be interesting to apply it on real customers with real data.

REFERENCE:

[1] A. Sahai, V. Machiraju, and K. Wurster, "Monitoring and controlling Internet based e-services," in *Internet Applications, 2001. WIAPP 2001. Proceedings. The Second IEEE Workshop on*, 2001, pp. 41-48.

[2] P. UL, "e-Service: Where is the Return on Investment," Gavle University, 2006.

[3] R. Allard and O. Hans, "Extending electronic portals with new services: exploring the usefulness of brand Extension models," *Journal of Retailing and Consumer Services*, vol. 12, pp. 245-254, 2005.

[4] R. K. Chellappa and S. Shivendu, "A model of advertiser—portal contracts: Personalization strategies under privacy concerns," *Information Technology and Management*, vol. 7, pp. 7-19, 2006.

[5] M. Hildebrandt. (2006) Profiling: From Data to Knowledge The challenges of a crucial technology. *Datenschutz und Datensicherheit*. 548-552.

[6] A. Corallo, G. Lorenzo, G. Solazzo, and D. Arnone, "Knowledge-Based Tools for E-Service Profiling and Mining," in *Mobile Information Systems*, B. Pernici, Ed., ed: Springer Berlin Heidelberg, 2006, pp. 265-291.

[7] W. Kangning, H. Jinghua, and F. Shaohong, "A Survey of E-Commerce Recommender Systems," in *Service Systems and Service Management, 2007 International Conference on*, 2007, pp. 1-5.

[8] B. Custers, "Effects of Unreliable Group Profiling by Means of Data Mining," in *Discovery Science*. vol. 2843, G. Grieser, Y. Tanaka, and A. Yamamoto, Eds., ed: Springer Berlin / Heidelberg, 2003, pp. 291-296.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9] H. ElGibreen and A. Mirza, "Group Profiling in E-Service Portals," presented at the EEE'11 - 10th Int'l Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government, USA, 2011.

[10] H. Evanschitzky, G. R. Iyer, C. F. Hofacker, R. E. Goldsmith, E. Bridges, and E. Swilley, "E-Services: A Synthesis and Research Agenda," in *E-Services*, ed: DUV, 2007, pp. 13-44.

[11] T.-H. Chou and J.-L. Seng, "An intelligent multi-agent e-services method—An international telecommunication example," *Information & Management*, vol. 46, pp. 342-350, 2009.

[12] J. P. Gant and D. B. Gant, "Web portal functionality and state government e-service," in *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, 2002, pp. 1627-1636.

[13] L. Hong-Chang, L. Tae-Ho, P. Yang-Su, and L. Myung-Joon, "CoPortlet: A WebDAV-based collaborative portlet supporting individualized services," in *Strategic Technologies, 2008. IFOST 2008. Third International Forum on*, 2008, pp. 289-292.

[14] J. Lu, D. Ruan, and G. Zhang, "E-Service Intelligence: An Introduction," in *E-Service Intelligence*. vol. 37, J. Lu, G. Zhang, and D. Ruan, Eds., ed: Springer Berlin / Heidelberg, 2007, pp. 1-33.

[15] A. Jameson and B. Smyth, "Recommendation to Groups," in *The Adaptive Web*. vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., ed: Springer Berlin / Heidelberg, 2007, pp. 596-627.

[16] M. Hildebrandt, "Profiling and AmI," in *The Future of Identity in the Information Society*, K. Rannenberg, D. Royer, and A. Deuker, Eds., ed: Springer Berlin Heidelberg, 2009, pp. 273-310.

[17] A. Freitas, *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Berlin: Spinger-Verlag, 2002.

[18] L. Deneubourg, S. Goss, N. Franks, C. Detrain, and L. Chretien, "The Dynamics of Collective Sorting: Robot-Like Ant and Ant-Like Robot," in *from animals to animats : proc. of the first int. conf. on simulation of adaptive behavior*, 1991, pp. 356 - 365.

[19] S. Wei, L. Jian-Chang, H. Yu-Jun, and L. Jian-Qiang, "Application of neural network model combining information entropy and ant colony clustering theory for short-term load forecasting," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, 2005, pp. 4645-4650 Vol. 8.

[20] T. Dinh and A. Mamun, "A combination of clustering algorithms with Ant Colony Optimization for large clustered Euclidean Travelling Salesman Problem," in *WSEAS SOSM 2004*, Miami, Florida, 2004, pp. 484-352.

[21] K. Taha and R. Elmasri, "SPGProfile: Speak Group Profile," *Information Systems*, vol. 35, pp. 774-790, 2010.

[22] K. Haewoon, S. Hwa-yong, Y. Jong-il, and M. Sue, "Connecting Users with Similar Interests Across Multiple Web Services," in *Third International ICWSM Conference*, 2009, pp. 246-249.

[23] M. F. Gomes and A. M. Canuto, "Carcara: A Multi-agent System for Web Mining Using Adjustable User Profile and Dynamic Grouping," in *Intelligent Agent Technology, 2006. IAT '06. IEEE/WIC/ACM International Conference on*, 2006, pp. 187-190.

[24] P. Kostkova, G. Diallo, and G. Jawaheer, "User profiling for semantic browsing in medical digital libraries," presented at the Proceedings of the 5th European semantic web conference on The semantic web: research and applications, Tenerife, Canary Islands, Spain, 2008.

[25] M. Julashokri, M. Fathian, and M. R. Gholamian, "Improving customer's profile in recommender systems using time context and group preferences," in *Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on*, 2010, pp. 125-129.

[26] C. Senot, D. Kostadinov, M. Bouzid, J. Picault, A. Aghasaryan, and C. Bernier, "Analysis of Strategies for Building Group Profiles," in *User Modeling, Adaptation, and Personalization*. vol. 6075, P. De Bra, A. Kobsa, and D. Chin, Eds., ed: Springer Berlin / Heidelberg, 2010, pp. 40-51.

[27] C. Ivan, C. Pablo, and V. David, "Enriching Group Profiles with Ontologies for Knowledge-Driven Colalborative Content Retrieval," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06. 15th IEEE International Workshops on*, 2006, pp. 358-363.

[28] R. Silva, V. Menezes, J. Oliveira, M. F. de Souza, C. de Mello, J. M. Souza, and G. Zimbrao, "Mining and analyzing organizational social networks for collaborative design," in *Computer Supported Cooperative Work in Design, 2009. CSCWD*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2009. 13th International Conference on, 2009, pp. 504-509.

[29] L. Vrdoljak, I. Bojic, V. Podobnik, G. Jezic, and M. Kusek, "Group-Oriented Services: A Shift towards Consumer-Managed Relationships in the Telecom Industry," in *Transactions on Computational Collective Intelligence II*. vol. 6450, N. Nguyen and R. Kowalczyk, Eds., ed: Springer Berlin / Heidelberg, 2010, pp. 70-89.

[30] H. J and K. M, *Data Mining: Concepts and Techniques*: Morgan Kaufmann, 2006.

[31] D. Trung and M. Abdullah, "A combination of clustering algorithms with Ant Colony Optimization for large clustered Euclidean Travelling Salesman Problem," in *WSEAS SOSM 2004*, Miami, Florida, 2004, pp. 484-352.