CHAPTER 5

5. EXISTING PERSONALIZATION APPROACHES

5.1 Introduction

The Personalization Process

Personalization aims to provide users with what they need without requiring them to ask for it explicitly. This means that a personalization system must somehow infer what the user requires based on either previous or current interactions with the user. This system obtains information on the user and infers what his needs are based on this information.

Web Personalization

The content on the Web in various fields is rapidly increasing and the need for identifying and retrieving the content exactly based on the needs of the users is more than required. Therefore, an ultimate need nowadays is that of predicting the user needs in order to improve the usability of Web information.

In brief, Web Personalization can be defined as any action that adapts the information or services provided by web to an individual user, or a set of users, based on knowledge acquired by their navigational behavior, recorded in the web logs. This information is often combined with the content and the structure of the site as well as the user’s interests/preferences.

Using the above specified sources of information as input to pattern discovery techniques, the system molds the provided content to the needs of each visitor of the web. The personalization process can result in the dynamic generation of suggestions, the creation of pages according to the needs of the user, highlighting of existing hyperlinks that are exactly required by the users. Most of the earlier research efforts in Web Personalization deal with Web Usage Mining [24].
Pure usage-based personalization, however, presents certain shortcomings, such as when there is insufficient use of data available in order to extract patterns, or when the web site's content changes and new pages are added but are not yet included in the web logs. The user’s visits usually aim at finding information concerning a particular subject, thus the underlying content semantics should be a dominant factor in the process of web personalization. There have been a number of research studies that integrate the web site’s content in order to enhance the Web Personalization process [25]. Most of these efforts characterize web content by extracting features from the web pages. Usually these features are keywords subsequently used to retrieve similarly characterized content based on the requirements of the user. When Web Personalization approaches were embedded with Semantic Web, it yields more effective search response and user satisfaction.

Web Mining is a mining of Web data on the World Wide Web. Web Mining does the process on personalizing these Web data. The Web data may be of the following.

- **Content** data are presented to the end-user appropriately structured. They can be simple text, images or structured data, such as information retrieved from databases.

- **Structure** data represent the way content is organized. They can be either data entities used within a Web page, such as HTML or XML tags, or data entities used to put a Web site together, such as hyperlinks connecting one page to another.

- **Usage** data represent a Web site’s usage, such as visitor’s IP address, time and date of access, complete path (files or directories) accessed, referrer’s address, and other attributes that can be included in a Web access log.
• *User profile* data provide information about the users of a Web site. A user profile contains demographic information (such as name, age, country, marital status, education, interests etc) for each user of a Web site, as well as information about user’s interests and preferences. Such information is acquired through registration forms or questionnaires, or can be inferred by analysing Web usage logs.

5.2 Existing Web Personalization Approaches

There are various existing approaches available to personalize the web data to enhance the user convenience. Some of the approaches are listed below.

*Content-based filtering*

These systems are solely based on individual user’s preferences. The system tracks each user’s behaviour and recommends them items that are similar to items the user liked in the past.

*Collaborative filtering*

These systems invite users to rate objects or divulge their preferences and interests and then return information that is predicted to be of interest for them. This is based on the assumption that users with similar behaviour (for example users that rate similar objects) have analogous interests.

*Rule-based filtering*

This is the system in which the users are asked to answer to a set of questions. These questions are derived from a decision tree, so as the user proceeds on answering them, what the user finally receives as a result (for example a list of products) is tailored to their needs. Content-based, rule-based and collaborative filtering may also be used in combination, for deducing more accurate conclusions.
**Clustering Based Approaches**

These approaches involve two approaches. These are item-based and user-based clustering. In user based clustering, users are clustered based on the similarity of their ratings of items (I). In item based clustering, items are clustered based on the similarity of ratings by all users in U. In the case of user-based clustering, each cluster centre $C_k^{(U)}$ is represented by an n-dimensional vector, $C_k^{(U)} = (ar_1, ar_2, ..., ar_n)$, where each $ar_j$ is the average item rating for (or average weight associated with) item $i_j$ by users in cluster $k$. In the case of item-based clustering the cluster centre is represented by an m-dimensional vector $C_k^{(I)} = (q_1, q_2, ..., q_m)$, where each $q_i$ is the average ratings by user, $u_i$, of items within the cluster.

**Web Personalization and User Profile**

The Web users expect more intelligent systems to gather the useful information from the huge size of Web to meet their information needs. The user profiles are created for user background knowledge description [26] [27]. User profiles represent the concept models possessed by users when gathering web information. A concept model is implicitly possessed by users and is generated from their background knowledge. This knowledge is used to gather relevant information about a user's preference and choices.

A user profile is a collection of personal data associated to a specific user. A profile refers therefore to the explicit digital representation of a person’s identity. Thus the user profile can be used to store the description of the characteristics of person. A user profile can also be described as the computer representation of a user model.

There are many models that have been developed for representing user profiles. These models provide knowledge from either a global or local knowledge
base. The *global analysis* uses existing global knowledge bases and to produce effective performance. The commonly used knowledge bases include generic ontology such as Word net, Thesauruses, Digital Libraries. The *local analysis* observes user behavior in user profiles. The user background knowledge can be better discovered and represented if global and local analysis is integrated. Local analysis is used for analyzing the user behavior in user profiles. It can be better improved by using ontological user profiles.

**Techniques using User Profiles**

![Diagram of user profiling methods](image)

**Figure 5.1: Methods of User Profiling**

The most common way to use a profile is to store information that enables personalization on an individual basis as represented in figure 5.1A. This is called Content-based Filtering which, applied to a textual document, evaluates the document's relevance by matching the keywords contained in a user profile with the keywords extracted from the text. On the Web, to prevent the user profiles transmitting through the network, user profiles are stored at the server.

Social or collaborative filtering is an effective way to take advantage of user profiles. This method collects the user profiles of a group of people and generates
recommendations based on the similarities of the profiles as given in the figure 5.1B. To implement collaborative filtering, the profiles of all users must be compared and therefore the best storage location is also to centralize them at the server.

A user profile can also be shared between different personalized applications that require the same user profile's content as in 5.1C. This collaboration enables both applications to gain a much more knowledge about the user's interest. Because all the personalized Web applications (on different servers) need to have access to the complete set of profiles for a specific user, it is required to store user profile at the browser.

**Management of User Profiles**

Several architectures are used for personalized services on the Web and they differ mainly in the locations of the management and storage functions.

**Figure 5.2: Various Architectures of Web Personalization**

The most common architecture is the server-based architecture, in which the user profiles are both stored and managed at the server as in figure 5.2A. Since the profiles of all users are centralized, the server needs to identify the user in order to extract the right user profile. This is done by using an authentication mechanism. This
architecture is efficient in that the user profiles do not transit through the network. The centralization of all the user profiles enables the use of both content-based and collaborative filtering but prevents user profiles from being shared between applications on different servers.

With this architecture, the service provider has to supply both hardware and software for the management and storage of the user’s profiles. For a worldwide service, those profiles may represent a large amount of data. The second architecture stores the user profiles on the client side and manages them on server side as given in the figure 5.2B. This architecture enables the use of content based filtering and profile sharing but not of collaborative filtering. The browser must provide a mechanism for permanently storing data on the user's computer, and this is a sensitive issue because most browsers, for security purpose do not allow a Web application (for example a Java applet) to permanently store any information on the terminal. The “Cookie” mechanism introduced by Netscape is an exception to this rule. By setting a cookie, an application can get data permanently stored by the browser and automatically sent back when the user accesses the application again. The main advantage of this second architecture is the distributed nature of the storage, which frees the service provider from supplying software and disk space for the database, but the transmission of the user profile between its storage location (client) and the management location (server) increases the response delay.

The third architecture manages and stores the user profiles on the client side as represented in figure 5.2C. In this, the personalization is done by the browser, and the architecture is therefore not a client-server architecture anymore (at least with respect to the personalization). This architecture enables the use of content-based filtering and user profile sharing but not of collaborative filtering. Although all these architectures
enable the use of content-based filtering, none of them can at the same time support collaborative filtering and the sharing of user profiles among different applications. Furthermore, no standard such as the Common Gateway Interface (CGI) has been defined for the management and storage of the user profiles on the server side. Each personalized Web application that uses the server-based architecture has to interface individually with the database that contains the user profiles.

**Semantic based Personalized Search**

Personalization aims to find a subset of Web data that matches the interest profile of a user or a group of users. This can be achieved by recommending Web pages or Websites to the users, or by filtering Web pages that are of interest to the users [28]. For example, this can done by analyzing the historical data recording user accesses to Web data, and mining the topics relevant to a user by clustering previously accessed Web pages based on content similarities. When a new Web page is found to be similar to one of the clusters, it can be routed to the user. Personalized search takes advantage of Semantic Web standards (RDF and OWL) to represent the content and the user profiles. Semantic based Personalization of Web data access can be effectively used for improving the precision and recall in search, particularly by re-ranking the search results based on the learner's past activities. The core part of Semantic approach on Web Personalization is the use of Ontology. As Web pages are annotated with ontology entity labels, the Web pages accessed by a user can lead to more effective content recommendation.

**Web Personalization and Ontology**

Ontology is used to represent user profiles in personalized web information gathering. Ontologies are the structural frameworks for organizing information. In computer science and information science, ontology formally represents knowledge as
a set of concepts within a domain, and the relationships between those concepts. It can be used to reason about the entities within that domain and may be used to describe the domain.

**Advantages of Ontology model in User Profile**

- An Ontology model discovered user background knowledge from user local instance repositories, rather than documents read and judged by users.
- Compared to the web data used by the web model, the Ontology model were controlled and contained less uncertainties.
- As large number of uncertainties eliminated when user background knowledge was discovered, as a result, the user profiles acquired by the Ontology model performed better than the web model.

**Personalized Ontology**

Personalized ontologies [29] are a conceptualization model that formally describes and specifies user background knowledge. Web users might have different expectations for the same search query. For example, for the topic “Apple”, an IT person may demand different information from normal users. An IT person expects “Apple” as system but normal users consider this as fruit. Sometimes even the same user may have different expectations for the same search query if applied in a different situation. Based on this observation, an assumption is formed that web users have a personal concept model for their information needs. A user’s concept model may change according to different information needs.

**Other Web Personalization Approaches**

Lot of research had been conducted in Personalized Ontology. Generally, personalization methodologies are divided into two complementary processes which
are (1) the user information collection, used to describe the user interests and (2) the inference of the gathered data to predict the closest content to the user expectation.

Information semantics are also used to enrich the personalization process, queries can be enriched by adding new properties from the available domain ontologies. The user modelling based on ontology can be coupled with dynamic update of user profile using results of information-filtering and Web usage mining techniques. Statistics collected through search engines show that spatial information is pervasive on the Web and that many queries contain spatial specifications, but it is more difficult to find relevant resources responding to query including a spatial component [30]. The spatial information personalization should consider spatial properties and relationships found in Web documents. Design of spatial Web applications requires at least three components: (1) a user model and associated user preference elicitation mechanisms and (2) a personalization engine combining spatial and semantic criteria and (3) a user interface enriched with spatial components. The spatial Web personalization requires the representation of user features, particularly those relevant to the spatial domain.

Middleton et.al. [31] explore the use of ontologies in the user profiling process within collaborative filtering systems. This work focuses on recommending academic research papers to academic staff of a University. The authors represent the acquired user profiles using terms of a research paper ontology (is-a hierarchy). Research papers are also classified using ontological classes. In this hybrid recommender system which is based on collaborative and content-based recommendation techniques, the content is characterized with ontology terms, using document classifiers (therefore a manual labeling of the training set is needed) and the ontology is again used for making generalizations/specializations of the user profiles.
Kearney and Anand [32] use an ontology to calculate the impact of different ontology concepts on the users navigational behavior (selection of items). In his work, they suggest that these impact values can be used to more accurately determine distance between different users as well as between user preferences and other items on the web site, two basic operations carried out in content and collaborative filtering based recommendations. The similarity measure they employ is very similar to the Wu & Palmer similarity measure. This work focuses on the way these ontological profiles are created, rather than evaluating their impact in the recommendation process.

B. Bhaskara Rao et al. [33] deliver an approach that learns the user profile and constructs fuzzy net by calculating togetherness between concepts, documents or both. This work has been done in two phases. In the first phase, the fuzzy nets with enriched extended user profile have been constructed. In second phase, the rank of each document has been evaluated by using clustering algorithm.
5.3 Comparison of Existing Personalization Approaches

As the result of conclusion, Table 5.1 gives the behaviour and limitation of various existing personalization approaches.

Table 5.1: Comparison of Existing Personalization Approaches

<table>
<thead>
<tr>
<th>Profiling Approach</th>
<th>Behaviour</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content based Filtering</td>
<td>Solely based on individual user’s preferences. This system tracks each user’s behaviour and recommends items that are similar to items the user liked in the past</td>
<td>Beneficiary is an individual user</td>
</tr>
<tr>
<td>Collaborative Filtering</td>
<td>Invites set of users for their preferences and interests. This is based on the assumption that users with similar behaviour have analogous interests</td>
<td>Does not follow clustered approach in storing users preferences</td>
</tr>
<tr>
<td>Rule based Filtering</td>
<td>Asks the users to answer to a set of questions</td>
<td>This approach is not an automatic method to get the behaviour of users</td>
</tr>
<tr>
<td>Cluster based Approach</td>
<td>Involves two approaches such as item-based and user-based clustering. In user based clustering, users (U) are clustered based on the similarity of their ratings of items (I). In item based clustering, items are clustered based on the similarity of ratings by all users.</td>
<td>Efficient use of storage is required to get faster response.</td>
</tr>
</tbody>
</table>

Cluster based Web Personalization approach has the benefits of both user profiling and clustering approach. User profiling enables to store and analyze the behaviour of users. The clustering approach facilitates to retrieve the personalized content faster.
5.4 Conclusion

Content based filtering and collaborative filtering are having major focus on identifying user preferences on Web. Cluster based approaches enhance the performance of the personalization system. User profiling and user’s usage data profiling are the efficient personalization methodologies. Even though there are various personalization approaches available, the key dimension along which personalization systems are evaluated include user satisfaction, accuracy, coverage, utility, robustness, performance and scalability.