

# Web Personalization Techniques in Data Mining

**S.Kannan<sup>1</sup>, Dr.G.N.K. Suresh Babu<sup>2</sup>**

<sup>1</sup>Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India.  
kannanmca2006@yahoo.co.in

<sup>2</sup>Professor, Department of Computer Applications, Acharya Institute of Technology, Bangalore, India.  
[gnksureshbabu@gmail.com](mailto:gnksureshbabu@gmail.com)

---

**Abstract:** In this paper we present an overview of Web personalization process viewed as an application of data mining techniques. Web personalization is the process of customizing a web site to the needs of each specific user or set of users. Personalization of a web site may be performed by the provision of recommendations to the users, highlighting/adding links, creation of index pages, etc. The web personalization systems are mainly based on the exploitation of the navigational patterns of the web site's visitors. When a personalization system relies solely on usage-based results, however, valuable information conceptually related to what is finally recommended may be missed. Nowadays when users navigate in internet they find lots of types of communication scattered across diverse sites. Due to the lack of users' interests in advertisements, primarily caused by not matching their preferences, today's advertisements have a low rate of success. Recently, owing to semantic web generation, some companies started to use personalized marketing in communication as a way to turn the conversions around and thus increase customer retention and loyalty. The main purpose of this paper is how web personalization techniques can be used for customized marketing with using data mining.

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

---

## 1. INTRODUCTION

For the past few decades, popularity of the internet has grown to a great extent with nearly every person, young or old, using it for a variety of purposes. People use the internet to get information in areas of interest, do research related to work or study, get good deals for commodities or travel, increase awareness about their surroundings and the world, get latest news, etc. With each passing day, large amounts of informative web sites, web pages or web documents get added to the already huge collection. Any popular search engine returns thousands of related links to a search query. It has become difficult for users to get the most relevant information from this plethora of related information readily available. Users often spend considerable time browsing the web pages for getting the right information. If the users' intention and interest for browsing a web page can be identified, it will be easier to make available that area of information with higher priority. Web personalization can be defined as any action that adapts the information or services provided by a web site to an individual user, or a set of users, based on knowledge acquired by their navigational behaviour, recorded in the web site's logs. This information is often combined with the content and the structure of the web site as well as the user's interests/preferences, if they are available. Using the four aforementioned sources of information as input to pattern discovery techniques, the system tailors the provided content to the needs of each visitor of the web site. The personalization process can result in the dynamic generation of recommendations, the create on of index pages, the highlighting of existing hyperlinks, the publishing of targeted advertisements or emails, etc. In this paper we focus on personalization systems that aim at providing personalized recommendations to the web site's visitors. The problem of providing recommendations to the visitors of a web site has received a significant amount of attention in the related literature.

## 2. WEB PERSONALIZATION

The process of providing information that is related to user’s current page is known as web personalization. This information is usually displayed on the current page in the form of web page links. The idea behind web personalization is that the web page currently being browsed by a user indicates his/her interest in that topic and it is likely that the user would be interested in more similar information. For example, in case of ecommerce the related information could be about other similar products to those that the user is viewing or about products that other users who bought or viewed this product also bought. This example would also work for a research or target oriented web browsing. The key information that is required for suggesting these similar web pages comes from the knowledge of other users who have also visited the current page as well as other pages before and after this current page. In addition to other users’ browsing information, web personalization can also take advantage of the web page content, the structure of the web page or the user’s profile information. All these help in creating a focused and personalized web browsing experience for the user. Web mining is the process that helps to collect and analyse information such as users’ browsing pattern, content of the web pages, structure of the web site, etc. and aid in creating a personalized web browsing experience. Fig.1 represents the components of web personalization.

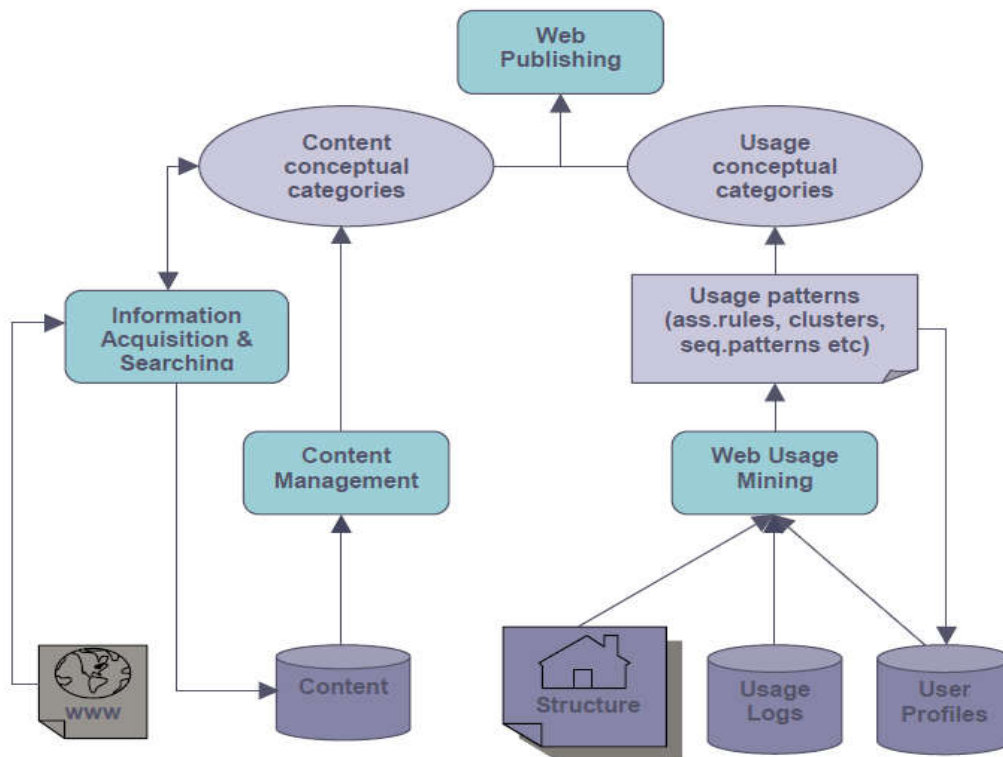


Fig.1 Components of Web Personalization

## 3. DIFFERENT PHASES OF WEB PERSONALISATION

Web personalization is the process of customizing a Web site to the needs of a specific user, considering the knowledge acquired from the analysis of the user’s navigational behaviour in to the account in correlation with other information collected in the form of structure, content, and user profile data. The web personalization process can be alienated into different phases namely collection of web data, pre-processing of web data, analysis of web data, and finally decision making or recommendation.

**3.1 Data Collection**

The process of gathering information either explicitly or implicitly specific to each visitor for recording their interests and behaviour while they browse a web site. The collection of activities completed in the past and recorded in Web server logs is known as the implicit data. This activity is performed by the web server and the user is not directly involved in collection of such data. The information submitted by the user at the time of registration or in response to the rating questionnaires is considered as the explicit data and it usually comes from the active involvement of the user. Explicit data collection requires users to exert most of the efforts. Web data in the form of content, structure, semantic, usage and user profile may be collected and used in the context of Web personalization. A user profile includes information about users’ interests and preferences and it contains demographic information for each user of a Web site.

**3.2 Pre-processing of data**

The data collected consists of various irrelevant information for example the log data collected from the web server are in the form of text files with a row for each http transaction. These data need to be cleaned before putting them for analysis. The Pre-processing task is performed to clean the data form inconsistencies. It filters out irrelevant information according to the goal of analysis

**3.3 Data analysis / Mining**

In this phase the specific data mining techniques that are popularly used for mining of web data or the machine learning techniques are applied to the processed data to discover interesting usage patterns. These usage patterns may form the groups according to the users’ behaviour. It classifies the content of a web site into semantic categories in order to make information retrieval and presentation easier for the user. This step is applied offline for automatic user profiling without adding the burden to the web server.

**3.4 Recommendation Phase**

This is the last phase deals with the actions that should be performed after taking the results of the previous analysis step. This phase usually performs the recommendations to the users by determining existing hyperlinks, the dynamic insertion of new hyperlinks that seems to be of interest for the current user to the last web page requested by the user, or even the creation of new index pages. Fig.2 represents phases of web personalization.

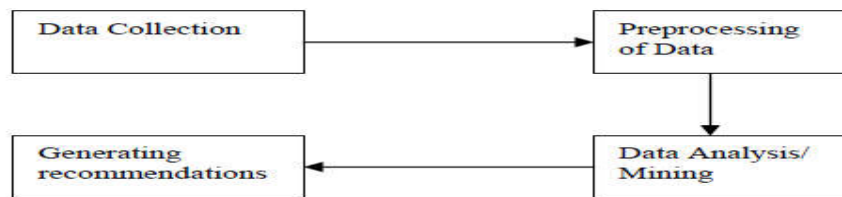


Fig.2 Phases of Web Personalization

**4. USER PERSONALIZATION**

In order to personalize a Web site, the system should be able to distinguish between different users or groups of users. This process is called user profiling and its objective is the creation of an information base that contains the preferences, characteristics, and activities of the users. In the Web domain and especially in e-commerce, user profiling has been developed significantly because Internet technologies provide easier means of collecting information about the users of a Web site, which in the case of e-business sites are potential customers. A user profile can be either static, when the information it contains is never or rarely altered, or dynamic when the user profile’s data change frequently. Such information is obtained either explicitly, using online registration forms and questionnaires resulting in static user profiles, or implicitly, by recording the navigational behaviour and/or the preferences of each user, resulting in dynamic user profiles.

In the latter case, there are two further options: either regarding each user as a member of a group and creating aggregate user profiles, or addressing any changes to each user individually. When addressing the users as a group, the method used is the creation of aggregate user profiles based on rules and patterns extracted by applying Web usage mining techniques to Web server logs. Using this knowledge, the Web site can be appropriately customized. Web Usage Mining techniques can be used to provide personalized web user experience. For instance, it is possible to anticipate, in real time, the user behaviour by comparing the current navigation pattern with typical patterns which were extracted from past web log. In this area, recommendation systems are the most common application; their aim is to recommend interesting links to products which could be interesting to users. Personalized Site Maps are an example of recommendation system for links.

## 5. WEB PERSONALIZATION TECHNIQUES

We describe three approaches used for generating a personalized Web experience for a user.

### 5.1 Content-Based Filtering

Content based filtering systems have their roots in information retrieval. The approach to recommendation generation is based around the analysis of items previously rated by a user and generating a profile for a user based on the content descriptions of these items. The profile is then used to predict a rating for previously unseen items and those deemed as being potentially interesting are presented to the user. A number of the early recommender systems were based on content-based filtering including Personal Web-Watcher, Info Finder, Newsreaders, Letizia and Syskill and Webert. Mladenic provides a survey of the commonly used text-learning techniques in the context of content filtering, with particular focus on representation, feature selection and learning algorithms. Syskill and Webert learns a profile from previously ranked Web pages on a particular topic to distinguish between interesting and non-interesting Web pages. To learn the profile, it uses the 128 most informative words, defined using expected information gain, from a page and trains a naïve Bayes classifier to predict future, unseen pages as potentially interesting or not for the user. The user may provide an initial profile for a topic, which in the case of Syskill and Webert, requires the definition of conditional probabilities for each word, given a page that is (not) interesting to the user. As pages get rated, these initial probabilities are updated, using conjugate priors, to reflect the rating of the pages by the user.

### 5.2 Traditional Collaborative Filtering

Goldberg et al. first introduced collaborative filtering as an alternative to content based filtering of a stream of electronic documents. The basic idea as presented by Goldberg et al. was that people collaborate to help each other perform filtering by recording their reactions to e-mails in the form of annotations. The application of this technology for recommending products has gained popularity and commercial success. Users provide feedback on the items that they consume, in the form of ratings. To recommend items to the active user, previous feedback is used to find other likeminded users. These are users that have provided similar feedback to a large number of the items that have been consumed by users. Items that have been consumed by likeminded users but not by the current user are candidates for recommendation. The assumption made by these systems is that users that have had common interests in the past, defined by feedback on items consumed, will have similar tastes in the future.

### 5.3 Model Based Techniques

Model based collaborative filtering techniques use a two stage process for recommendation generation. The first stage is carried out offline, where user behavioural data collected during previous interactions is mined and an explicit model generated for use in future online interactions. The second stage, is carried out in real-time as a new visitor begins an interaction with the Web site. Data from the current user session is scored using the models generated offline, and recommendations generated based on this scoring. The application of these models are generally computationally inexpensive compared to memory-based approaches such as traditional collaborative filtering, aiding scalability of the real-time component of the recommender system.

Model generation can be applied to explicitly and implicitly obtain user behavioural data. While the most commonly used implicit data is Web usage data, data pertaining to the structure and content are also often used?

## 6. WEB PERSONALIZATION CHALLENGES

Personalization has gone through different development phases since early 2000. It initially started as a tool to attract and retain visitors by giving them chance to explore more of the site. Advertising and promoting products and services, nevertheless, were part of this phase as well. The next phase attempted to increase turnover of customer's spending by offering more expensive or similar products. Today, personalization is increasingly used as a means to speed up the delivery of the right information to a visitor in order to customize products and services for meeting and exceeding his or her requirements. Those companies that are systematically gathering information about their customers, product attributes, purchase contexts and integrate it with behavioural segmentation such as demographics, attitudes and buying patterns can make more sophisticated offers that identify customers who are most likely to defect. This personalization strategy ultimately increases number of regular customers and amount of each transaction and has made personalization as a required and expected feature of an e-business.

For example, Timberland Boot Studio by allowing customer to select different leathers and colours gets three times hits on its customized boots. Without challenging the enormous potential and contribution of personalization technology, the question still remains whether personalization really works? The smart answer is it depends. On one hand the benefits could be significant not only for the website visitor (being offered more interesting, useful and relevant web experience) but also for the provider. On the other hand, personalization requires extensive and precise data that are neither easily obtainable, nor can be mined and analysed efficiently. As such in many cases the output does not seem so successful in understanding and satisfying user needs and goals. First and foremost, the ethical dimension of personalization need to be addressed, since online user's navigations are recorded for building and updating user profiles and this can put privacy of users in jeopardy. At the same time users are becoming more vigilant and have higher expectation. They are not so happy with idea of being stereotyped without their consent. Users also expect to be treated equal and have enough control and choices.

The cost of technology initiated from intelligent software and storing hardware as well as the time spent are also critical factors which must be justified in the long run. Schneider and Bowen proposed to explain that customer satisfaction of the service originates with the handling of three basic customer needs: security, justice and self-esteem. The term personalization refers to the customized flow of communication that sends different recipients marketing tailored to their individual preferences. In order to tailor these messages organizations have to evaluate what consumer information, such as demographics, psychographics or purchase history, to use in order to be personal in their communication. The digital technology has empowered companies to personalize and customize messages in order to communicate with stakeholders. This has started a development of direct marketing and reshaped the way companies target and segment markets, create dialogues and challenge the old approach to mass marketing. However, personalization today is a sensitive area and often twinned with privacy issues. The degree of personalization will vary over time depending on where in the customer lifecycle the customer is.

The more intimate the relation is, the more positive the consumers will respond to personalization. A benefit of customized ads is that they are tailored to the needs of the consumer and the high fit often enhances the consumer's intention to purchase. But there are two sides of the coin. A message with high fit may also reveal that the sender has used information about the consumer, which the consumer may recognize as a loss of control. The author argue that the new ways that companies use personalization have created new privacy challenges. These trends in personalization is; social-based personalization, behavioural profiling and location-based personalization. Social-based personalization has to do with the growth of social media. Facebook, Twitter and other social media have access to lots of information about their users such as real name, demographics, personal photos and current location.

This information is used in various situations like, when people search for friends, receiving targeted advertising or other personalized recommendations. These are the most used forms of social-based personalization. This form of personalization is easy to use for social networks since people are often very willing to share their information. Behavioural profiling is the practice of collecting data about users' activities online through a system. These systems track users over a long period of time using browser cookies and often have little or no permission from the users, compared to social networks. The assumption that is made when using profiling is that users will like the same things that similar users do. The system creates clusters of users who share similar characteristics, based on their past preferences. Location-based personalization on the other hand only offers personalized ads based on your location.

The author conducted a study where they examined how consumers reacted to personalized e-mails. Their findings showed that an increased knowledge of effective personalization may help companies to increase click-through rates, and just as important knowledge of what cause negative response. Negative response from consumers on personalized e-mails may cause much harm to the brand or company and it is important to eliminate that type of marketing. Overly restrictive marketing and other limited offerings may jeopardize future business or end an existing consumer-company relationship. To avoid this White et al recommend companies to have great knowledge about their audiences and target the right segment from the beginning so that they will find the personalization useful for them and Like all forms of communication it is important to understand the recipient's behaviour and communication via e-mail enables a high degree of personalization.

## 7. FUTURE TRENDS IN WEB PERSONALIZATION

The Web is an incubator for a large spectrum of applications involving user interaction. User preferences and expectations, together with usage patterns, form the basis for personalization. Enabling technologies include data mining, pre-processing, sequence discovery, real time processing, scalable warehousing, document classification, user modelling and quality evaluation models. As websites become larger, more competitive, and more dynamic, and as users become more numerous, and more demanding, and as their interests evolve, there is a crucial need for research that targets the above enabling technologies, and leads them toward the path of scalable, real-time, online, accurate, and truly adaptive performance. From another perspective, the inherent and increasing heterogeneity of the Web has required Web based applications to integrate a variety of types of data from a variety of channels and sources. The development and application of Web mining techniques in the context of Web content, usage, and structure data will lead to tangible improvements in many Web applications, from search engines and Web agents to Web analytics and personalization. Future efforts, investigating architectures and algorithms that can exploit and enable a more effective integration and mining of content, usage, and structure data from different sources promise to lead to the next generation of intelligent Web applications.

## 8. CONCLUSION

In this paper, we have provided a comprehensive review of trends and techniques for Web personalization. Personalization from web mining has been received lots of interests in business as a gifted tool to improve sales and retain customers, since, it can increases customer's satisfaction by providing them with tailor made products and services. Web Personalization can also help the company to implement build-to-order policy by connecting customer's requirements and preferences directly to production line and supply chain. We have taken the view that Web personalization is an application of data mining and must therefore be supported during the various phases of a typical data mining cycle. We have described the various explicit and implicit data sources available along with the typical approaches used to transform this data into useful user profiles/models that can be used to generate recommendations. We have also described various approaches to generating recommendations from a set of user profiles/ models. Research into this topic has raised a number of interesting issues related to the personalization process.

These are issues that need to be addressed by any personalization system that aims to provide robust, accurate and useful personalized content to its users. We also provide a description of the current understanding of how these systems should be evaluated, describing some of the most commonly used metrics within personalization literature. Because of the explosive proliferation of the Web, Web personalization has recently gained a big share of attention, and significant strides have already been accomplished to achieve WWW personalization while facing tough challenges. However, even in this slowly maturing area, some newly identified challenges beg for increased efforts in developing scalable and accurate web mining and personalization models that can stand up to huge, possibly noisy, and highly dynamic web activity data. Along with some crucial challenges, we have also pointed to some possible future direction in the area of WWW personalization.

## REFERENCES

- [1] Agrawal R. and Srikant R. (2000). *Privacy-preserving data mining*, In *Proc. of the ACM SIGMOD Conference on Management of Data*, Dallas, Texas, 439-450.
- [2] Bettina Berendt. *Using site semantics to analyze, visualize, and support navigation*. *Data Mining and Knowledge Discovery*, 6(1):37–59, 2002.
- [3] Bettina Berendt, Bamshad Mobasher, Miki Nakagawa, and Myra Spiliopoulou. *The impact of site structure and user environment on session reconstruction in web usage analysis*. In *Proceedings of the 4th WebKDD 2002 Workshop, at the ACM-SIGKDD Conference on Knowledge Discovery in Databases (KDD'2002)*, 2002.
- [4] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. *Effective personalization based on association rule discovery from web usage data*. In *Web Information and Data Management*, pages 9–15, 2001.
- [5] Catherine Bounsaythip and Esa Rinta-Runsala. *Overview of data mining for customer behavior modeling*. Technical Report TTE1-2001-18, VTT Information Technology, 2001.
- [6] Eirinaki M., Vazirgiannis M. (2003). *Web mining for web personalization*. *ACM Transactions On Internet Technology (TOIT)*, 3(1), 1-27.
- [7] I. Varlamis, M. Vazirgiannis, M. Halkidi, B. Nguyen, THESUS, *A Closer View on Web Content Management Enhanced with Link Semantics*, in *IEEE Transactions on Knowledge and Data Engineering Journal*, June 2004 (Vol. 16, No. 6), pp. 585-600.
- [8] Joachims T. (2002). *Optimizing search engines using clickthrough data*. In *Proc. of the 8th ACM SIGKDD Conference*, 133-142.
- [9] Mobasher B., Dai H., Luo T., and Nakagawa M. (2001). *Effective personalization based on association rule discovery from Web usage data*, *ACM Workshop on Web information and data management*, Atlanta, GA.
- [10] Mobasher, B.: *Web usage mining*. In Wong, J., ed.: *Encyclopedia of Data Warehousing and Data Mining*. Idea Group Publishing (2005) 1216–1220
- [11] Mobasher, B.: *Web usage mining and personalization*. In Singh, M.P., ed.: *Practical Handbook of Internet Computing*. CRC Press (2005)
- [12] M. Eirinaki, *New Approaches to Web Personalization*, PhD Thesis, Athens University of Economics and Business, Dept. of Informatics, 2006

- [13] M. Eirinaki, C. Lamos, S. Pavlakis, M. Vazirgiannis, *Web Personalization Integrating Content Semantics and Navigational Patterns*, in *Proc. of the 6th ACM International Workshop on Web Information and Data Management (WIDM'04)*, November 2004, Washington DC
- [14] Nasraoui O., Krishnapuram R., and Joshi A. (1999). *Mining Web Access Logs Using a Relational Clustering Algorithm Based on a Robust Estimator*, 8th International World Wide Web Conference, Toronto, 40-41.
- [15] Kardan, A., Fani, M.R. and Mohammadian, N. (2011), *Purposing an Architecture for Learner Modeling base on Web Usage Mining in e-Learning Environment*, The 5th Data Mining Conference Dec. 14 - 15, 2011 ; Tehran, Iran
- [16] Lam, S., Riedl, J.: *Shilling recommender systems for fun and profit*. In: *Proceedings of the 13th international conference on World Wide Web*. (2004) 393–402
- [17] Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: *Item-based collaborative filtering recommendation algorithms*. In: *Proceedings of the 10th International World Wide Web Conference*. (2001)
- [18] S. Paulakis, C. Lamos, M. Eirinaki, M. Vazirgiannis, *SEWeP: A Web Mining System supporting Semantic Personalization*, in *Proc. of the ECML/PKDD 2004 Conference, Pisa, Italy, September 2004*
- [19] Tan, P., Kumar, V., Srivastava, J.: *Selecting the right objective measure for association analysis*. *Information Systems* (2004) 293–313
- [20] Veeramalai Sankaradass, Kannan Arputharaj “An Intelligent Recommendation System for Web User Personalization with Fuzzy Temporal Association Rules” *European Journal of Scientific Research* ISSN 1450-216X Vol.51 No.1 (2011), pp.88-96 © EuroJournals Publishing, Inc. 2011