

Web Usage Mining for Personalization

Dipti N. Belsare
Department of MCA
PIET
Nagpur

Amit A. Dabli
2nd author's affiliation
Module Lead
Ness Technology Pvt.Ltd, Pune

B. V. Ambulkar
Department of MCA
GHRIIT
Nagpur

ABSTRACT

This paper is about study of personalization in terms of web usage mining and ecommerce. In the first part we briefly introduce the web mining and its types the content web mining, structure web mining and the usage mining. Then we discuss the different aspects of personalization in detail. To conclude, we present some emerging trends and some future perspectives.

General Terms

This paper is focusing on the terms like data mining, web mining. Categories of web mining web content mining, web structure mining and Usage mining. And web personalization.

Keywords

Web mining, web usage mining, personalization.

1. INTRODUCTION

Web mining approaches provide insight for achieving web personalization. The works contributed towards personalization of a website. Web usage patterns could be deciphered by applying data mining techniques on web usage records of the surfers. These techniques include Association Rule Mining, Clustering, Classification and Sequential Patterns Mining. The personalization of services offered by a Web site is an important step in the direction of alleviating information overload, making the Web a friendlier environment for its individual user and hence creating trustworthy relationships between the Website and the visitor-customer. In [1] Web Personalization is simply defined as the task of making Web-based information systems adaptive to the needs and interests of individual users. Typically, a personalized Website recognizes its users, collects information about their preferences and adapts its services, in order to match the user's needs. Web personalization improves the Web experience of a visitor by presenting the information that the visitor wants to see in the appropriate manner and at the appropriate time. In e-business, Web personalization additionally provides mechanisms to learn more about customer needs identify future trends and eventually increase customer loyalty to the provided service.

2. WEB MINING

Web mining is broadly defined as the discovery and analysis of useful information from the WWW. Target data sets for data mining in the context of the web are classified into the following types:

2.1 Web Content Mining

Web content mining is the process of extracting knowledge from the content of web documents. Its goal is gathering data

and identifying patterns related to the contents of the web and the searches performed on them. There are two main strategies: Web page mining, extracting patterns directly from the contents existing in web pages. In this case the data in use

2.2 Web Structure Mining

Web structure mining is the process of inferring knowledge from the structure of data. This specialty intends to reveal the real structure of web sites through the gathering of structure related data, and mainly about its connectivity. Typically it takes into account two types of links: static and dynamic. [12]

2.3 Web usage Mining

WUM is the process of discovering and interpreting patterns of user access to the web information system. Web Usage Mining applications are based on data collected from three main sources [17]: (i) web servers, (ii) proxy servers, and (iii) web clients.

The primary data sources used in Web usage mining are the server log files, which include Web server access logs and application server logs.

The aims of web usage mining

- The discovered patterns are usually represented as collections of pages or items that are frequently accessed by groups of users with common needs or interests.
- Such patterns can be used to better understand behavioral characteristics of visitors or user segments,
- Improve the organization and structure of the site, and create a personalized experience for visitors by providing dynamic recommendations.

Most sophisticated systems and techniques in usage mining are parsed into three distinctive processes: pre-processing, pattern discovery, and pattern analysis [18]. Every process can be categorized as follows see figure1.

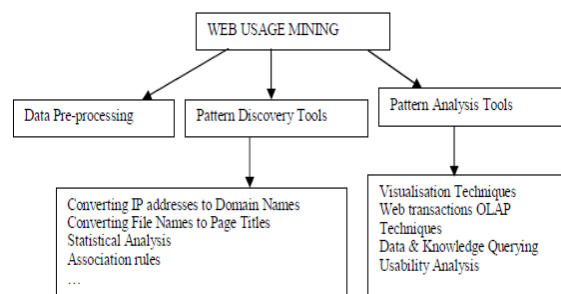


Fig 1: Web usage mining process categories

3. WEB USAGE MINING AND PERSONILIZATION

The work on Web usage mining can be a source of ideas and solutions towards realizing Web personalization. Usage data,

such as those that can be collected when a user browses a specific Web site, represent the interaction between the user and that particular Web site. Web usage mining provides an

approach to the collection and preprocessing of those data, and constructs models representing the behavior and the interests of users. These models can be used by a personalization system automatically; hence, a Web personalization system can employ Web usage mining methods in order to achieve the required robustness and flexibility. Web usage mining consists of the basic data mining stages:

. **Data Collection-** During this stage, usage data from various sources are gathered and their content and structure is identified. For Web usage mining, data are collected from Web servers, from clients that connect to a server, or from intermediary sources such as proxy servers and packet sniffers. A number of techniques that have been employed at this stage, can be used to attain efficient collection of user data for personalization.

. **Data Preprocessing-** This is the stage where data are cleaned from noise, their inconsistencies are resolved, and they are integrated and consolidated, in order to be used as input to the next stage of Pattern Discovery. In Web usage mining, this involves primarily data filtering, user identification and user session identification. The techniques that are used here can provide efficient data elaboration.

. **Pattern Discovery-** In this stage, knowledge is discovered by applying machine learning and statistical techniques, such as clustering, classification, association discovery, and sequential pattern discovery to the data. The patterns required for Web personalization, correspond to the behavior and interests of users. This is the stage where the learning methods are applied in order to automate the construction of user models.

. **Knowledge Post-Processing-** In this last stage, the extracted knowledge is evaluated and usually presented in a form that is understandable to humans, e.g. using reports, or visualization techniques. For Web personalization the extracted knowledge is incorporated in a Personalization module in order to facilitate the personalization functions.

Figure 1 summarizes graphically the above-described stages.

4. PERSONILIZATION

Personalization is a process of gathering and storing information about site transactions, analyzing the information, and, based on the analysis, delivering the more relevant information to people who visit Web site. A number of personalization techniques can enable more customized messages for customers in advertisements, product recommendations, news feeds and other marketing tactics. Providing personalization for real-time applications can affect system performance; therefore, the deployment of

personalization is important and should be integrated into the overall system design. This is especially true for high volume Web sites. Web site type should determine selection of personalization techniques. Regardless of type, Web sites look increasingly to the use of personalization to increase repeat business. Web personalization can be described as any action that makes the Web experience of a user customized to the user's taste or preferences. Principal elements of Web personalization include modeling of Web objects (such as pages or products) and subjects (such as users or customers), categorization of objects and subjects, matching between and across objects and/or subjects, and determination of the set of actions to be recommended for personalization. [1]

To-date, the approaches and techniques used in Web personalization can be categorized into three general groups: manual decision rule systems, content-based filtering agents, and collaborative filtering systems. [1]

A Web personalization system can offer a variety of functions. Four basic classes of personalization functions are: memorization, guidance, customization and task performance support [13]. Each of these is examined in more detail below.

Memorization: - This is the simplest form of personalization function, where the system records and stores in its 'memory' information about the user, such as name and browsing history. When the user returns to the site, this information is used as a reminder of the user's past behavior, without further processing. Memorization is usually not offered as a stand-alone function, but as part of a more complete personalization solution.

Guidance:- Guidance as a personalization function refers to the endeavor of the personalization system to assist the user in getting quickly to the information that the user is seeking in a site, as well as to provide the user with alternative browsing options. This personalization function not only increases the users' loyalty but also alleviates in a great extent the information overload problem that the users of a large Web site may face.

Customization:- Customization as a personalization function refers to the modification of the Web page in terms of content, structure and layout, in order to take into account the user's knowledge, preferences and interests. The main goal is the management of the information load, through the facilitation of the user's interaction with the site.

Task Performance Support:-

Task performance support is a functionality that involves the execution of a particular action on behalf of a user. This is the most advanced personalization function, inherited from a category of Adaptive Systems known as personal assistants, which can be considered as client-side personalization systems. The same functionality can be envisaged for the personalization system employed by a Web server.

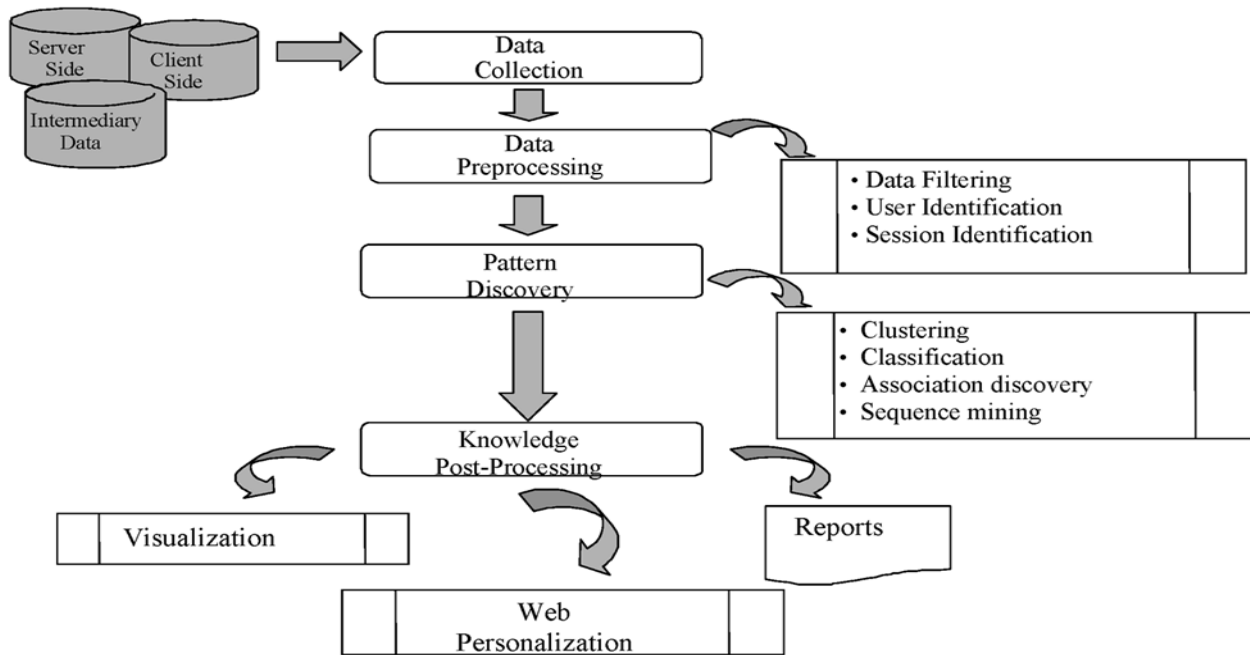


Fig 2: Web usage mining process

5. APPROACHES TO WEB PERSONALIZATION

During the evolution of the Web, personalization has been recognized as a remedy to the information overload problem and as a means of increasing visitor loyalty to a Web site.

Due to the importance of personalization for Web-based services, several Web personalization techniques have been proposed in the past few years. Although it is not in the scope of the survey to present these techniques in detail, a brief overview of the approaches is presented below.

Web personalization techniques classified into three generic approaches:

(a) Manual decision rule systems- According to this approach, a Web-based service is personalized via manual intervention of its designer and usually with the cooperation of the user. Typically, static user models are obtained through a user registration procedure and a number of rules are specified manually concerning the Web content that is provided to users with different models. Two examples from a wide range of products that adopt this approach are Yahoo!'s personalization engine and Websphere Personalization (IBM).

(b) Content-based filtering systems- This group of techniques applies machine learning methods to Web content, primarily text, in order to discover the personal preferences of a user. A tool that adopts this approach is News Weeder, which is able to adaptively construct user models from a user's browsing behavior, based on the similarity between Web documents containing news items. These models can be used to filter news items according to each user's requirements.

(c) Social or collaborative filtering systems- The aim of this approach is to personalize a service, without requiring the analysis of Web content. Personalization is achieved by searching for common features in the preferences of different users, which are usually expressed explicitly by them, in the form of item ratings, and are recorded by the system. The Recommendation Engine (Net Perceptions) and Websphere Personalization (IBM) are examples of products that use also

this method, while its most renowned application is in the amazon.com electronic shop.

6. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the subject presented herein.

7. CONCLUSION AND FUTURE DIRECTIONS

Personalization requires analysis of business goals and the development of business requirements, use cases, and metrics.

Once these are fully understood, organizations may find that their personalization strategies don't require substantial augmentation of their application environments.

Organizations should also examine how their needs are likely to change and whether their approach to personalization will enable them to move along the continuum from preliminary targeting efforts to tapping the potential of every possible segment.

8. REFERENCES

- [1] Web Usage Mining and Personalization Bamshad Mobasher, DePaul University
- [2] Analysis of Web Usage Mining Hui Yu, Zhongmin Lu School of Management, South -Central University for Nationalities, Wuhan, P.R.China
- [3] Intelligent Techniques for Web Personalization Sarabjot Singh Anand and Bamshad Mobasher
- [4] Web Mining – Accomplishments & Future Directions Jaideep Srivastava, Prasanna Desikan, Vipin Kumar
- [5] Web Mining by Juan C. Dürsteler
- [6] Efficient and Anonymous Web-Usage Mining for Web Personalization Cyrus Shahabi, Farnoush Banaei-Kashani

- [7] Web site personalization IBM developer works
- [8] Web Usage Mining & Personalization Dr. Guandong Xu,
- [9] Intelligent Web & Information Systems (IWIS) Department of Computer Science, Aalborg University
- [10] A Conceptual Model for Website Personalization and Web Personalization Kavita Das and O.P. Vyas
- [11] Intelligent Techniques for Web Personalization Sarabjot Singh Anand and Bamshad Mobasher
- [12] How to Win Online: Advanced Personalization in E-Commerce An Oracle White Paper March 2011
- [13] A Near Real-Time Personalization for eCommerce Platform Amit Rustagi
- [14] Web Usage Mining as a Tool for Personalization: A Survey Dimitrios Pierrakos, Georgios paliouras, Christos Papatheodorou and Constantine D. Spyropoulos
- [15] R. Cooley, B. Mobasher, and J. Srivastava. Data Preparation for Mining World Wide Web Browsing Patterns. Journal of Knowledge and Information Systems, 1(1):5–32, 1999.
- [16] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations, 1(2):12–23, 2000
- [17] Web usage mining: Discovery and applications of usage patterns from web data. Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan SIGKDD Explorations, 1(2):12–23, 2000.
- [18] Web Mining, <http://www.galeas.de/webmining.html>.