

A SURVEY ON WEB MINING AND DIFFERENT APPROACHES FOR WEB PERSONALIZATION

Kinjal P. Patel¹, Asst. Prof. Vidhu Singhal²

¹M.E. Student, Department of Computer Science and Engineering

² Asst. Prof, Department of Computer Science and Engineering
SPB College of Engineering, Linch, Gujarat, India.

Abstract: The information on the web is growing dramatically, without a help system; the users may spend lots of time on the web finding the information they are interested in. Web mining is used to discover and extract useful information from the World Wide Web. The main goal of web mining technique is to find user's access object automatically from the large amount of web log data. Web mining has been explored to different techniques Web structure mining, Web content mining and Web usage mining that have been used for the variety of the application. Web usage mining techniques used to discover usage pattern from web data, in order to understand and better serve the needs of web based applications. In preprocessing, pages on a web sites are processed to create proper structure of a web site. In usage mining, we count the support of all the data in the web log file to find frequent itemsets based on URL related analysis and classify the web pages in to two category i.e either its index page or content page. In Site recommendation phase, the web site is examined to find better ways to give proper navigate system to the user. Web Personalization is the capability to customize customer communication based on knowledge preferences and behaviours at the time of interaction. Here we proposed a new approach, which generate personalization process more accurate and less time consuming.

Keywords: Personalized recommendation, Web usage mining, Data preparation, Web log.

I. INTRODUCTION

Web mining is an appliance of data mining techniques to large web log data repositories. The tremendous growth in the number and the complexity of information resources and services on the Web has made Web personalization an indispensable tool for both Web-based organizations and for the end users. Web personalization can be defined as any action that personalizes the Web experience to a particular user, or a set of users [1]. The ability of a site to engage visitors at a deeper level, and to successfully guide them to useful and pertinent information, is now viewed as one of the key factors in the site's ultimate success. 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 modelling 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. [2]

Web mining is the application of data mining techniques to extract knowledge from web data, in the form of web content, web structure, and web usage mining [3].

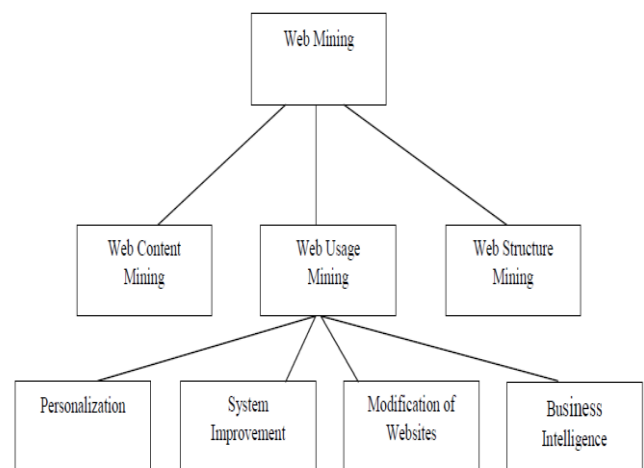


Fig.1.Taxonomy of Web Mining

Web Content Mining: Is the process of extracting useful information from the contents of Web documents. Content data corresponds to the collection off acts a Web page was designed to convey to the users. It may consist of text, images, audio, video, or structured records such as lists and tables.

Web Structure Mining: The structure of a typical Web graph consists of Web pages as nodes, and hyperlinks as edges connecting between two related pages. In addition, the content within a Web page can also be organized in a tree structured format, based on the various HTML and XML tags within the page. Thus, Web Structure Mining can be regarded as the process of discovering structure information from the Web. This type of mining can be performed either at the (intra-page) document level or at the (inter-page) hyperlink level.

Web Usage Mining: Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Some of the typical usage data collected at a Web site include IP addresses, page references, and access time of the users.

II. CONCEPT OF WEB USAGE MINING

Discovery of meaningful patterns from data generated by client-server transactions on one or more Web servers.

Typical Sources of Data:

- a. Automatically generated data stored in server.
- b. Access logs, referrer logs, agent logs, and client-side cookies.
- c. E-commerce and product-oriented user events.
- d. (E.g. shopping cart changes, ad or product click-throughs, etc.).
- e. User profiles and/or user ratings.
- f. Meta-data, page attributes, page content, site structure.

There are three main tasks for performing WUM— pre-processing, pattern discovery and pattern analysis. As below [4]:-

Pre-processing: It is generally used as groundwork of data mining practice, data pre-processing cleaned/filtered the raw data to eliminate outliers or irrelevant items, grouping individual page accesses into semantic units for the purpose of the user. The different types of pre-processing in Web Usage Mining are - usage, content, and structure pre-processing.

The pre-processing phase of HTTP server information is made up of the following five steps:

- a. Data cleaning
- b. User identification
- c. User session identification
- d. Path completion (creation of user session file)
- e. Transaction Identification (creation of transaction file)
- f. transaction file)

Pattern Discovery: In this, Web Usage Mining can be able to unearth patterns in server logs and carried out only on samples of data. Interpretation and evaluation of results be done on samples of data. The various pattern discovery methods are— Statistical Analysis, Association Rules, Clustering, Classification, Sequential Patterns, and Dependency Modeling.

Pattern Analysis: The need behind pattern analysis is to filter out uninteresting rules or patterns from the set found in the pattern discovery phase. Most common form of pattern

analysis consists of a knowledge query mechanism such as SQL. Content and structure information can be used to filter out patterns containing pages of a certain usage type, content type, or pages that match a certain hyperlink structure.

AREA OF WEB USAGE MINING

- a. Personalization
- b. System Improvement
- c. Site Modification
- d. Business Intelligent
- e. Usage Characterization

REQUIREMENTS OF WEB USAGE MINING

- a. Gather useful usage data,
- b. Filter out irrelevant usage data,
- c. Establish the actual usage data,
- d. Discover interesting navigation patterns,
- e. Display the navigation patterns clearly,
- f. Analyze and interpret the navigation patterns, correctly, and
- g. Apply the mining results effectively.

III. PERSONALIZATION ON THE WEB MINING

Web personalization is a strategy, a marketing tool, and an art. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in your content delivery framework to manipulate what information you present to your users and how you present it [5]. Web personalization is defined as any action that adapts the information or services provided by a Web site to the needs of a particular user or a set of users, taking advantage of the knowledge gained from the users' navigational. It is necessary to stress the difference between layout customization and personalization. In customization the site can be adjusted to each user's preferences regarding its structure and presentation. In personalization systems modifications concerning the content or even the structure of a Web site are performed dynamically [6, 7].

The new recommendation system is developed with the following processes: (i) Collection of web data (ii) Data preparation (also termed as data preprocessing in web usage mining field), (iii) finding frequent itemsets, (iv) web page classification, and (v) Web page recommendation.

1. Collection of Web data

Implicit data includes past activities/clickstreams as recorded in Web server logs and/or via cookies or session tracking modules. Explicit data usually comes from registration forms and rating questionnaires. Additional data such as demographic and application data (for example, e-commerce transactions) can also be used. In some cases, Web content, structure, and application data can be added as additional sources of data, to shed more light on the next stages.

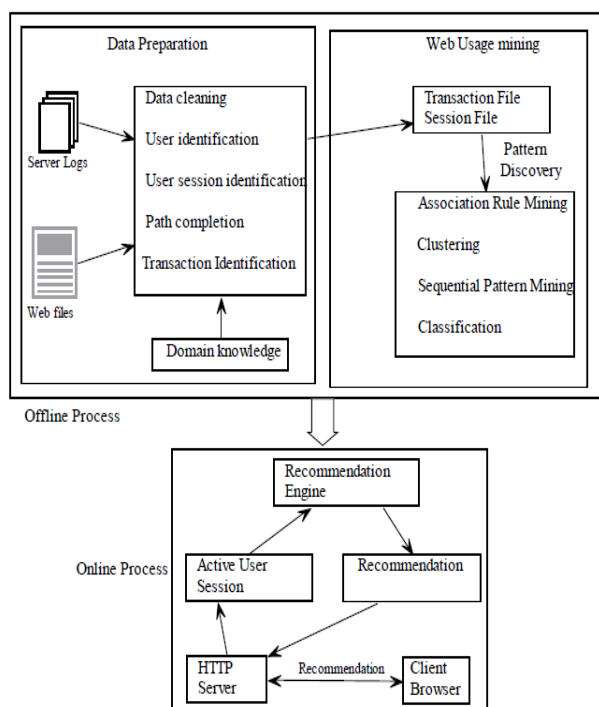


Fig.2. Usage based Personalized Recommendation System

2. Preprocessing of Web data

Data is frequently pre-processed to put it into a format that is compatible with the analysis technique to be used in the next step. Preprocessing may include cleaning data of inconsistencies, filtering out irrelevant information according to the goal of analysis (example: automatically generated requests to embedded graphics will be recorded in web server logs, even though they add little information about user interests), and completing the missing links (due to caching) in incomplete clickthrough paths. Most importantly, unique sessions need to be identified from the different requests, based on a heuristic, such as requests originating from an identical IP address within a given time period.

3. Finding frequent itemsets

For the Usage mining phase, we will measure the parameter like support and confidence for the different section from the web site. In this phase, we count the support of all the data in the web log file.

4. Classification and Analysis of Web data

In this phase, we divide us pages in to two categories. i) Index page and ii) content page. The Pages which is used by the user for the navigation of the system is called the index page. The Content page is a page containing information in which the user interested. Also known as Web Usage Mining [8], this step applies machine learning or Data Mining techniques to discover interesting usage patterns and statistical correlations

between web pages and user groups. This step frequently results in automatic user profiling [9], and is typically applied offline, so that it does not add a burden on the web server.

5. Decision making/Final Recommendation Phase

The last phase in personalization makes use of the results of the previous analysis step to deliver recommendations to the user. The recommendation process typically involves generating dynamic Web content on the fly, such as adding hyperlinks to the last web page requested by the user. This can be accomplished using a variety of Web technology options such as CGI programming.

IV. CONCLUSION

The Web is providing a direct communication medium between the vendors of products and services, and their clients. Web usage mining model is a kind of mining to server logs. Web Usage Mining plays an important role in realizing enhancing the usability of the website design, the improvement of customers' relations and improving the requirement of system performance and so on. Web usage mining provides the support for the web site design, providing personalization server and other business making decision, etc.

REFERENCES

- [1] Manoj Swami, Prof. Manasi Kulkarni "Understanding Web Personalization With Web Usage Mining and its Application: Recommender System" (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 5, May 2013.
- [2] Dilpreet kaur, Sukhpreet Kaur "A Study on User Future Request Prediction Methods Using Web Usage Mining" International Journal of Computational Engineering Research, Vo. 03, Issue, 4.
- [3] Schafer J.B., Konstan J., and Reidel J. (1999). "Recommender Systems in Ecommerce", In Proc. ACM Conf. E-commerce, 158-166.
- [4] Gopal Pandey, Swati Patel "A Process Oriented Perception of Personalization Techniques in Web Mining" ISSN: 2319-6386, Volume-1, Issue-2,
- [5] A.Jebaraj Ratnakumar "An Implementation Of Web Personalization Using Web Mining Techniques" Journal Of Theoretical And Applied Information Technology
- [6] Mobasher, B., Web Usage Mining and Personalization, in Practical Handbook of Internet Computing, M.P. Singh, Editor. 2004, CRC Press. p. 15.1-37.
- [7] Schafer J.B., Konstan J., and Reidel J. (1999). Recommender Systems in ECommerce, In Proc. ACM Conf. E commerce, 158-166. Nasraoui O., Krishnapuram R., and Joshi A. (1999). Mining Web Access Logs Using a

- [8] Relational Clustering Algorithm Based on a Robust Estimator, 8th International World Wide Web Conference, Toronto, 40-41.
Srivastava, J., Cooley, R., Deshpande, M., And
- [9] Tan, P-N. (2000). Web usage mining: Discovery and applications of usage patterns from web data, SIGKDD Explorations, 1(2), 12-23.

I.