A Survey on Web Transactions: Application, Process and Techniques

[1] Pushpraj Singh Chauhan, [2] Dr. Asheesh Shah [3] Dr. Suresh Jain
[1]Research Scholar, [2]Professor, [3]Director
[1, 2] Mewar University [3] PIEMR Indore

Abstract: Due to the tremendous growth in the World Wide Web, there has been large amount of web data generated. This web data can be mined to provide results which can greatly improve the performance and help business to grow. In this paper, we have discussed about the type of web mining which uses the web server logs as data sources known as web usage mining. We have surveyed on web usage mining in terms of different applications, Process and techniques which includes clustering, classification, association rule mining etc. We also studied about the various applications of web usage mining.

Index Terms **- Association Rule Mining, Clustering, Pattern discovery, Web Mining, Web Transaction.**

I. INTRODUCTION

The Internet and the World Wide Web (WWW) have made dramatic impact on individuals and organizations in less than a decade. E-commerce has been on a steady rise. However, transaction processing on the web is not the dominant use of the Internet or e-commerce although it is an essential application. Some transactions are very simple, such as purchasing a book or transferring funds, and can be processed immediately. Other transactions are simply defined, transaction processing is the unambiguous and independent execution of a set of operations on data in a database, which treats the set of actions as a single event [1].

Web transaction

Web transaction or just a transaction is a sequence of URLs combined into one complete process. Typical web transactions are when a customer logs in a member website, makes a purchase on a shopping site, fills in and submits a web form and performs other interactions with a website and web application.

The web provides different type of information that can be discovered using mining techniques:

1. Web user activity, from server logs and Web browser activity tracking.
2. Web graph structure, from links between pages, people and other data.
3. Web content, for the data found on Web pages and inside of documents.

A sample web log record is shown below:

123.46.7.79.8 - [12/Mar/2017:04:06:50 -0500] ―GET/HTTP/1.0‖ 200 3240

Where,

- 123.46.7.79.8- IP address
- "-"(hyphen) indicates Anonymous user id
- 12/Mar/2017:04:06:50- Web page access time
- -0500- The time zone
- GET/HTTP- HTTP request method
- 200- HTTP status code
- 3240- Number of bytes transmitted.

II. WEB MINING

1. Extracting knowledge from the content of documents or their descriptions is known as web content mining.
2. Web structure mining means inferring knowledge from the World Wide Web organization andlinks between references.
3. Finally, web usage mining, also known as Web Log Mining, is the process of extracting interesting patterns in web access logs.

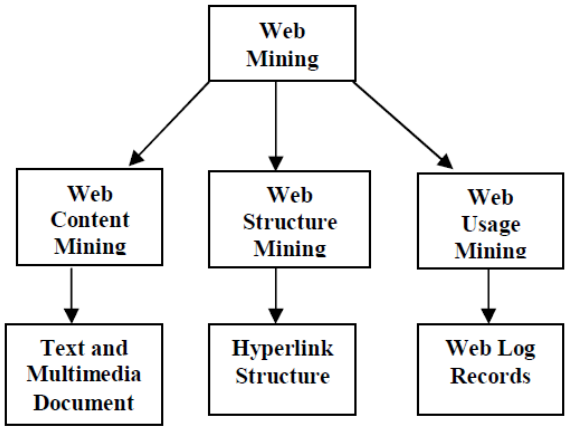

Fig. 1: Web Mining Categories

Web Usage Mining

Web usage mining tries to discover the useful information from the secondary data derived from the interactions of the users while surfing on the Web. It focuses on the techniques that could predict user behavior while the user interacts with Web.

Web usage mining is the application of data mining techniques to discover interesting usage patterns from web usage data, in order to understand and better serve the needs of web-based applications[2] [3]. A web user whenever interacts with the web server leaves behind traces of information which when analyzed helps improving the design of the system. Web servers record these user interactions which captures the identity or origin of web users with their browsing behavior in the access logs. The three main tasks for performing Web Usage Mining are Data Gathering, Preprocessing, Pattern Discovery and Analysis. Preprocessing consists of converting the usage, content and structure information contained in the various available data sources into the data abstraction necessary for pattern discovery [3].

III. APPLICATIONS

Web Usage Mining deals with study of Web transactions. It also deals with the Study of Web Logs. The results produced by the mining of web logs can be used for various purposes. There are various applications for web usage mining in different areas, and such applications are:

A. To Personalize the Web content

Web Usage Mining Techniques are used to personalize the Web Pages, depending on user's Web Transactions. Personalization plays an important role in making a stronger bonding, to build marketing strategies which are acceptable to all , and to regulate the promotion of products for customers. Also, web transactions can be used to obtain information that supports website designers [11].

B. To Improve System

Web usage mining is used to improve the performance of web servers and web-based applications. By understanding the web transactions produced by the web users various policies and strategies related to the web traffic, can be produced for web caching, network transmission, load balancing and data distribution [7].

C. To Provide Security

By using web usage mining we can obtain various patterns from web log which can be used in intrusion detection, attempted break-ins, fraud, etc. [8].

D. To Provide Site Design Support

Usability plays an important role in the design and implementation of websites. Web usage mining provides the results about the user's behavior which helps the designers in redesigning the content and structure of the website. Besides, some tools are also there which can be used to automatically change the structure of the web site based on usage patterns discovered [8].

E. Enhancing E-learning environment

Various tools are available which can be used for extraction the user's information from the web log. These extracted information can be used for personalizing the web pages and for tracking the activities of the user on the internet. For example, designers can identify the web pages that are always visited, web pages that are never visited, and the group of users that visit specific web sites [12].

F. Business Intelligence

The business intelligence is used to improve companies performances and to maintain competitive advantage in the market place, i.e., by using business intelligence companies can make the best decisions quickly and efficiently. Web usage mining is the appropriate technique for extracting information and building a useful and knowledgeable database about customer behaviors. Web Usage Mining also helps in determining different marketing strategies, i.e., it can also be used to increase the sales and place the company's products on a higher level [5].

IV. WEB USAGE MINING PROCESS

a) Data Gathering

Most of the web usage mining systems use the log data as their primary source of data. A Web log file records activity information when a Web user submits a request to a Web server. A log file can be located in three different places: i) Web servers, ii) Web proxy servers, and iii) client browsers.

b) Data Preprocessing:

Preprocessing consists of converting the usage, content and structure information contained in the various available data sources into the data abstraction necessary
for pattern discovery [3]. This step includes: i) data cleaning, ii) identifying user, iii) building sessions and
iv) path completion.

c) Pattern Discovery:

This step involves using statistical method to carry on the analysis and mine the processed data. We may discover the user or the user community's interests then construct interest model. At present the usually used machine
learning methods mainly have clustering, classifying, the relation discovery and the order model discovery.
Each method has its own excellence and shortcomings, but the
quite effective method mainly is classifying and clustering at the present.

d) Pattern analysis:

Challenges of Pattern Analysis are to filter uninteresting information and to visualize and interpret the interesting patterns to the user. Pattern analysis requires the knowledge of query languages such as SQL. The summarized data is loaded into a data cube and OLAP operations are performed to get better results.

V. WEB USAGE MINING TECHNIQUES

I. Pattern Discovery

Pattern discovery involves the employment of sophisticated techniques from artificial intelligence, data mining techniques, psychology and information theory in order to extract knowledge from collected and pre processed data. Some of the most widely used pattern discovery approaches are statistical analysis, association rule mining, clustering, classification, sequential patterns and dependency modeling [4]

II. Association Rules Mining

In web usage mining, those pages that are accessed together can be referred by association rules. Minimum support value can also be used in it, this in turn can be helpful in organizing web space quickly and easily. This can be understood as follows:
Suppose 85% of users who visited page A also visited page B. So, the web designer should put a direct link between the two pages [10]. In association mining the application mainly focuses on the prediction of the next most interesting page for the user [6]. Some of the algorithms for association mining are [13]:

1) Markov Chains
2) Improved Apriori All
3) Fp growth and Prefixspan
4) Custom-built Apriori algorithm

III. Clustering:

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [2]. In the Web Usage Mining, there are two types of interesting clusters to be discovered, usage clusters and page clusters. Clustering of users tends to find groups of users showing similar browsing patterns. Such knowledge is useful for inferring user demographics in order to perform market segmentation in E-commerce applications or provide personalized Web content to the users. Also clustering of pages will discover groups of pages having related content. This information is very useful for Internet search engines and Web assistance providers.

IV. Classification

Classification is function that assign items in a collection to target categories/ classes [9]. In web usage mining, a profile of users can be developed belonging to a specific class by using classification [8]. Example, classification can provide useful rules like: 50% of users who order online from website A are aged 25-35 and live in South West. Some of the algorithms which can be used for different purposes are given below [6]-[13]:

1. HCV, CDL4 are algorithms for extraction rules that represent user interests.
2. RIPPER, C4.5, Naïve Bayesian are algorithms for prediction of an interesting page.
3. Rough Theory Set is an algorithms for classification of sessions according to a concept.

V. Sequential Patterns

Sequential patterns discovery is to find the inter-transaction patterns such that the presence of a set of items is followed by another item in the time-stamp ordered transaction set. Web log files can record a set of transactions in time sequence. If the web-based companies can discover the sequential patterns of the visitors, the companies can predict users' visit patterns and target market on a group of users.

VI. Dependency modeling.

It consists of techniques that are aimed at finding a model describing dependencies between variables in the Web domain. This is potentially useful for predicting future Web resources consumption. For example, it could help develop strategies to increase the sales of products offered by Website [4]

VII. CONCLUSION

Web Mining is the process of extracting knowledge from the huge data repository the World Wide Web. In this paper, firstly we gave an introduction about the web mining and also studied the categories of web mining. Web Mining a part of data mining in terms of Web includes three categories, Web Content Mining, Web Structure Mining and Web Usage Mining. In the next sections, we focus on mainly Web Usage Mining and the process of web usage mining. Then we studied the different applications, process and techniques of web usage mining**.**

REFERENCES

[1] Pete, L., "Transaction Processing," Computerworld, Vol.35, No.40, 2001.
[2] Srivastava J., Cooley R., Deshpande, M. and Tan, P. N., Web Usage Mining: Discovery and applications of Usage Patterns from web data', SIGKDD Explorations 2000.
[3] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd Edition, Elsevier.
[4] J. Shrivastava, R. Cooley, Mukund Deshpande and Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns form Web Data," SIGKDD Explorations, Vol. 1, Issue 2, pp.12-23, Jan 2000.
[5] R. Cooley, B. Mobasher, and J. Srivastava, "Web mining: information and pattern discovery on the World Wide Web," in Proc. Ninth IEEE International Conference on Tools with Artificial Intelligence, 1997, pp. 558-567.
[6] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, "Web usage mining as a tool for personalization: A survey," User Modeling and User-Adapted Interaction, vol. 13, pp. 311-372, 2003
[7] F. M. Facca and P. L. Lanzi, "Mining interesting knowledge from weblogs: a survey," Data & Knowledge Engineering, vol. 53, pp. 225-241, June 2005.
[8] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," ACM SIGKDD, vol. 1, 2000.
[9] D. Dixit et al., "Mining access patterns using classification," International Journal of Engineering Science and Technology, vol. 2, 2010.
[10] K. M, R. Jadhav, R. J. D. Dixit, T. Khodkar, and A. Nehete, "Pattern discovery using association rules," International Journal of Advanced Computer Science and Applications, vol. 2, 2011.
[11] S. P. Nina, M. Rahman, K. I. Bhuiyan, and K. E. U. Ahmed, "Pattern discovery of web usage mining," in Proc. International Conference on Computer Technology and Development, 2009, pp. 499-503.
[12] O. R. Zaïane, "Web usage mining for a better web-based learning," presented at the Conference on Advanced Technology for Education, 2001.
[13] S. K. Pani, L. Panigrahy, V. H. Sankar, B. K. Ratha, A. K. Mandal, and S. K. Padhi, "Web usage mining: A survey on pattern extraction from web logs," International Journal of Instrumentation, Control & Automation (IJICA), vol. 1, 2011.