

Analysis to Improve Performance of Website Using Pattern Recognition Techniques

Neha Sharma

Gwalior Institute of Technology & Science, Gwalior M.P

Abstract: Web usage mining is a main research area in Web mining focused for learning about Web users and their interactions with Web sites to know the level of usage the website. The motive of mining is to find users' access models automatically and quickly from the vast Web log data, such as frequent access paths, frequent access page groups and user clustering. Through web usage mining, the server log, registration information and other relative information left by user access can be mined with the user access mode which will provide foundation for decision making of organizations. The main motive of this research is, to find a complete framework and findings in Web usage patterns from Web log files of a real Web site that has all the difficult aspects of real-life Web usage, including developing user profiles and external data describing database of the Web content.

Keywords: Web mining, Websites, Models, User log

1. Introduction

Information on Internet is increasing rapidly day by day, and it is applied for different purposes in decision support system. Web mining is the application of data mining and is used for the same task. Web mining is classified into three categories: Web Usage Mining (WUM), Web Content Mining (WCM), and Web Structure Mining (WSM). Among these, WUM is applied on usage data and it is being used at large scale by the organizations to study the behavior of their web users. In WUM the user's web log is collected for inferring the useful information by analyzing it. In present scenario every organization trusts on their websites for the growth of their business. The organizations collect the data from their web server to analyze the behavior and investigating interest of the users. The ability to track user browsing behavior down to individual mouse click has brought the vendor and end customer closer than ever before, it is now possible for a vendor to personalize his product message for individual customer.

Pattern recognition is the task of finding useful information from web server logs applying various techniques such as filtering, grouping etc. This extracted knowledge plays a very important role in formulation of important rules (decisions) regarding organization website structure, making marketing and advertising more fruitful and effective. Before the process of pattern recognition the log file data has to go through three stages i.e., preprocessing, pattern discovery and pattern analysis.

2. Related Work

In Web Mining, data can be collected at the server side, client-side, proxy servers, or obtained from an organization's database. Every type of data collection is different. The data collection is differs in terms of the location of the data source and also the type of data available, the segment of population from where the data was collected, and its method of implementation is also different.

There are many types of data that can be used in Web Mining. These can be considered as

Real Data: The real data also known as content . The data on the Web page that was designed to convey with the users. This usually consists of, but is not limited to, text and graphics.

Structure: Structural Data which describes the organization of the content. Intra-page structure information includes the arrangement of various HTML or XML tags within a given page. It can be represented structure like a tree, where the <html>tag becomes the root of the tree.

Usage: This type of data describes the pattern of usage of Web pages, such as IP addresses& page references, and the date and time on which the page is accesses.

Profile Data: Data that provides some information about users of the Web site. Includes registration data and customer profile information like login Id Password etc

3. Problem Formulation

In today's world data on Internet is increasing per day and the administrator's are continuously trying to make their website more friendly and efficient for their users. Pattern extracted from web server log helps them to make restructuring of websites and implementation of new applications which will increase their traffic and business. In this paper the problem defined is the extraction of patterns from web server log file.

In this paper, our aim is to take out experimental work on web log data collected from web server to find out useful browsing patterns. Results from the experiment will play an important role in improving the web server performance. There are a number of web usage mining tools available but here Web Log Explorer (WLE) tool is used for the implementation of our work . The uses of this application are,

It used to identify the usage level of one particular web site. The user's behavior on one particular website will be monitored by this application.

Through this application we can answer for the following six questions:

- 1) How are people using the particular web site?
- 2) Which pages are being accessed most frequently
- 3) To know which page (or) hyperlink got maximum number of hits.
- 4) Users details will be easily maintained.
- 5) With the help of the web log file details the reports will be produced.
- 6) With the results of the reports we can use web as a more profitable one for particular business strategy.

3.1 Pattern Analysis or Discovery

Pattern discovery techniques includes algorithms to discover patterns from web data. Once user transactions or sessions have been initiated, there are several types of access pattern mining that can be performed depending on the needs of the analyst. Some of them are discussed below

3.1.1 Sequential Pattern

Discovery of sequential patterns is to find the inter-transaction patterns such that the presence of a set of data is followed by another data in the time-stamp ordered transaction set. Web log files record a set of transactions in timely sequence to find out the pattern .

3.1.2. Group Analysis

This technique is useful to extract high level information. In this technique similar information is grouped .For e.g. group of all visitors receiving Unsuccessful Response Code (URC) from server. It means how many users are not able to get their request fulfilled. This kind of information helps in website performance as webmaster is able to filter out user's requesting non available web pages by looking out response code. In this paper, log file data is grouped according to the response code sent by the server. For grouping of the visitors based on the response code, log file used in not preprocessed.

3.1.3 Decision Trees

A decision tree is a flow chart of questions or data points that ultimately leads to a decision

3.1.4 Dynamic Site Analysis

Most of the Websites were created using static HTML pages. Today, a number of companies make systems that allow an HTML file that can be dynamically created around a database. This offers advantages like, including of centralization of storage, flexibility, and version control.

4. Pattern Analysis

Pattern analysis is also used in Web Usage mining process. The motivation behind pattern analysis is to filter out patterns which are worthless. The analysis method is usually governed by the application for which Web mining is done. The most common form of pattern analysis consists of a

knowledge query mechanism such as SQL. Another method is to load usage data into a data cube in order to perform operations.

5. Conclusion

We are trying to framework & study the web usage mining with pattern recognition techniques and carried out experimental work on data collected from use of a website to find out useful browsing patterns Pattern recognition can be used by the web administrator to optimize web site performance The extracted usage pattern will help administrators in managing the website resources in better way. The results or findings from this study are surely useful for web administrator in order to improve web site performance through the improvement contents, structure, presentation and delivery.

References

- [1] "Web Analytics An hour a day", Avinash kaushik, Wiley India Publication,2007.
- [2] "Web Analytics 2.0", Avinash kaushik, Wiley India Publication,2010.
- [3] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/The MIT Press, 1996.
- [4] J. Ross Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.
- [5] Resul Das, Ibrahim TURKOGLU, "Extraction of Interesting Patterns Through Association Rule Mining for Improvement of Website Usability," Journal of Electrical & Electronics Engineering, Vol. 9, No. 2, 2009.