

A Review on Analysis to Improve Performance of Website

Neha Sharma

Gwalior Institute of Technology & Science Gwalior M.P

Abstract: This paper will focus on Web usage mining for improvement in performance of a website. Web usage mining consists of three phases: Pre-processing, Pattern discovery and Pattern analysis. A detailed description will be given for each part of them, however, special attention will be paid to the user navigation patterns discovery and analysis. In this paper the user privacy is an another important issue. Web mining, will be included to make it easier to understand how the data mining techniques will apply to large Web data repositories to extract important information.

Keywords: Web mining, Pre-processing, Pattern discovery, Pattern Analysis

1. Introduction

The Web mining research relates to several research communities such as Database, Information Retrieval and Artificial Intelligence. Although there exists quite some confusion about the Web mining, the most recognized approach is to categorize Web mining into three areas: Web content mining, Web structure mining, and Web usage mining. Web content mining focuses on the discovery/retrieval of the useful information from the Web contents/data/documents, while the Web structure mining emphasizes to the discovery of how to model the underlying link structures of the Web. The distinction between these two categories isn't a very clear sometimes. Web usage mining is relative independent, but not isolated, category, which mainly describes the techniques that discover the user's usage pattern and try to predict the user's behaviors.

2. Data Pre -processing for Mining

Web usage mining is the application of data mining techniques to use logs (secondary Web data) of large Web data repositories. The results generated can be used in the design tasks such as Web site design, Web server designed and of navigating through a Web site. The data can be collected at the server-side, client-side, proxy servers, or obtained from database. For each type of data collection, the difference is not only the location, but also the available data type, the segment of population from which the data was collected and the method of implementation. The information sources available to mine include Web usage logs, Web page descriptions. The preprocess has three different conversions: Usage converting, Content converting, and Structure converting.

3. Diagram shows the High level Data processing

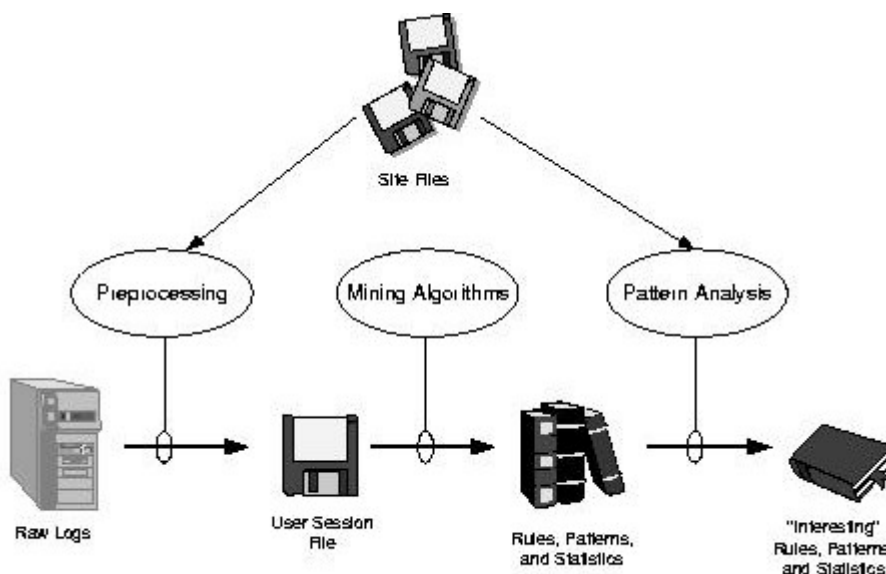


Figure 1: High Level Web Usage Mining Process

4. Pattern Discovery

Pattern discovery converges the algorithms and techniques from several research areas, such as data mining, machine learning, statistics and pattern recognition.

4.1 Statistical Analysis

Statistical techniques is a powerful tools for extracting knowledge about visitors of a Web site. The analysts may perform different kinds of descriptive statistical analyses based on different variables when analyzing the session file. By analyzing the statistical information contained in the periodic Web the result can be useful for improving the system performance, enhancing the security of the system, facilitation the site modification task, and providing support for marketing decisions.

4.2 Sequential Pattern

Sequential pattern technique is used to find the inter-session pattern, such that a set of the data items follows the presence of another's in a time-ordered set of sessions . It's very useful for the Web marketer to predict the future trend, which is useful for advertisements aimed at certain user groups. Sequential patterns also include some other types of temporal analysis such as trend analysis, change point detection, or similarity analysis.

5. Pattern Analysis

Pattern Analysis is a final stage of the whole Web usage mining. Main task of this process is to eliminate the irrelative rules or patterns and to get the useful rules or patterns from the output(result) of the pattern discovery process. The output of Web mining algorithms is often not in the form suitable for direct human use , so that we have to convert into format that can be used easily. This can be converted with help of analysis methodologies and tools. There are two most common approaches for the pattern analysis. One is to use the knowledge query mechanism such as SQL, while another is to construct multi-dimensional data cube before perform OLAP operations .

6. Conclusion

The area of research in this paper is Web mining with the focus on the Web Usage Mining. Three types of web data mining are introduced. Around the key topic of this paper - usage mining, we provide detailed description of the three phases of the process. These are data preprocessing pattern analysis & pattern Discovery.

References

- [1] B. Berendt. Web usage mining, site semantics, and the support of navigation
- [2] J. Borges and M. Levene. Data mining of user navigation patterns. In Proceedings of the WEBKDD'99 Workshop on Web Usage Analysis and User Profiling, August 15, 1999, San Diego, CA, USA, pages 31-39, 1999

- [3] R. Cooley, B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the world wide Web. In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997
- [4] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide Web browsing patterns. Knowledge and Information Systems, 1(1), 1999
- [5] R. Cooley. Web Usage Mining: Discovery and Application of Interesting Patterns from Web data. PhD thesis, Dept. of Computer Science, University of Minnesota, May 2000