WEB USAGE MINIG IN MODERN ERA

Abstract:

Web mining is the use of data mining techniques to automatically catch and abstract information from web. It helps to solve the problem of inventing how users are using a website. It involves mining logs and get meaning full data from web logs. It is the function of data mining techniques to web data where at least one of structure or usage data is used in the mining process. The present paper accord with primary discussion of web mining and also help to future research and enhanced from this paper.

Keywords:

Web mining, web content mining, web structure mining, web usage mining, hyperlink structure, Pattern Discovery, Pattern Analysis

Introduction:

Web mining is the operation of data mining techniques to come upon the patterns from the World Wide Web. As the name says, this is information gathered by mining the web. Web content, hyperlinks and server logs. The goal of Web mining is to look for patterns in Web data by collecting and analyzing information in order to gain insight into trends, the industry and users in general.

The contents of data mined from the Web may be a collection of facts that Web pages are meant to contain, and these may consist of text, structured data such as lists and tables, and even images, video and audio. The Web is the universal information space that can be accessed by companies, governments, universities, teachers, students, customers, businessmen and some users. In this universal space trading and advertising activities are held. No one knows the size of the World Wide Web.

Web mining is an area that lately has gained a lot of interest. There are three modes of Web mining according to the usage of Web data used as input in data mining process, namely Web Usage Mining (WUM), Web Content Mining (WCM) and Web Structure Mining (WSM) which are shown in fig. 1.
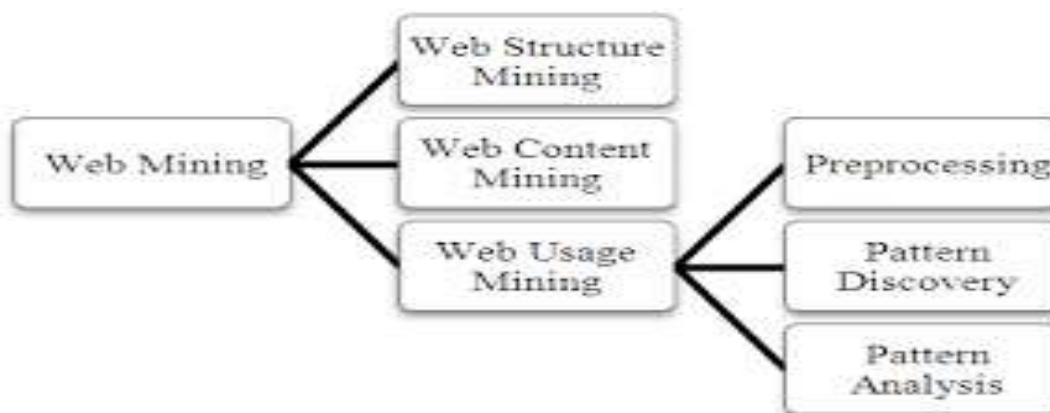
Area of Web Mining Can be Used
Examples:

- Web search, e.g. Google, Yahoo, MSN, Ask, ...
- Specialized search: e.g. Froogle (comparison shopping), job ads (Flipdog)
- eCommerce:
- Recommendations: e.g. Netflix, Amazon
- improving conversion rate: next best product to offer
- Advertising, e.g. Google AdSense
- Fraud detection: click fraud detection, ...
- Improving Web site design and performance

Web Structure Mining:

Web structure mining uses graph theory to consider the node and connection structure of a web site and one of its task is to identify more desirable documents. It is tries to discover the link structure of the hyperlink at the inter- document level.

Conforming to the type of web structural data, web structure mining can be part into two kinds:

1. **Abstracting patterns from hyperlinks in the web:** a hyperlink is a structural component that connects the web page to a different location.
2. **Mining the document structure**: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

The aim of Web structure mining is to generate structural summary about the Web site and Web page. If a Web page is link up to another Web page directly, or the Web pages are friends we would like to catch the relationships among those Web pages. The relations maybe enter one of the types, such as they related by equivalent or ontology, they may have similar contents, both of them may sit in the same Web server so that created by the same person. Another work of Web architecture mining is to discover the nature of the grouping or network of hyperlink in the Web sites of appropriate domain. This may help to checked the flow of information in Web sites that may represent some particular domain, therefore the query processing will be easier and more efficient.

Web Contenet Mining:

Web content mining objective is the knowledge discovery, in which the main objects are the traditional selection of multimedia documents such as images, video, and audio, which are embedded in or associated to the web pages.
Web content mining is clashing from the data mining and text mining. It is related to data mining due to it's usage of data mining approach in web content mining. And it is related to text mining because of most of the web content are texts. Web mining is semi structure essence of the web while text mining has unstructured texts.
Web content mining could be separated from two points of view:

- Agent-based way or Database way. The first way aims on developing the information discover and filtering.
- The second way aims on design the data on the Web into more structured form in order to affix standard database querying system and data mining applications to analyze it.

Web Usage Mining:

Web usage mining is the type of Web mining achtion that affect the automatic discovery of user access patterns from one or more Web servers. The Web Usage Mining is the application of data mining

technique to discover the useful patterns from web usage data. It can discover the user access patterns by mining log files and associated data of particular web site.

Web Usage Mining to be found in steps Data Collection, Pre-processing, Pattern Discovery and Pattern Analysis. In Web Usage Mining, data can be gathered from server log files that include web server access logs and application server logs. Data is also achieved from site files and working databases. The data collected in web log file is incomplete and not suitable for mining directly. Pre-processing is needed to change the data into suitable form for pattern detection

Pre-processing of web data is the process of conversion of the raw data into a usable data model. Pattern discovery step uses several data mining design is used to abstract the user patterns. Finally, pattern inquiry from web data uncovers useful and alluring user patterns and direction. These steps are normally executed after the web log data is together.

Three main methods that are used to web usage mining include: Association rules, sequential patterns, and clustering. In this section, each of these methods are described as below

- **Association Rules**

✓ Association rule is the most essential rule of data mining methods which is used more than other approach in the web usage mining. This rule shows, if user observes A and B pages, most likely will observe page C at the same meeting. A common algorithm to extract association rules is Apriori algorithm.

- **Sequential Patterns**

✓ Sequential patterns are used to discover the subsequence in the large volume of sequential data. In web usage mining, sequential patterns are used to find user navigation patterns which appear frequently at meeting

- **Clustering**

✓ Clustering techniques diagnose groups of similar items among high volumes of data. This is done based on distance functions which measures the degree of similarity between different items. Clustering in web usage mining is used for array of similar meetings.

- **Web Mining Benifical Areas**

✓ E-Learning

✓ Digital libraries

✓ E-Government

✓ Electronic commerce

✓ E-Politics and E-Democracy

✓ Security and Crime investigation

✓ Electronic Business

Conclusion:

Web and web usage continues is grow, so too grows the opportunity to analyses web data and extract all manner of useful knowledge from it.in this paper briefly described the web mining and applications of web mining some areas of future research. It can also be used to provide fast and efficient services for business.it is expected that more applications of web mining will be developed.

References

i. http://shodhganga.inflibnet.ac.in/bitstream/10603/44185/11/11_chapter%203.pdf

ii. https://www.researchgate.net/publication/236673264_Web_Usage_Mining_Process_and_Techniq uesJones, P. W. (1999)

iii. https://www.cs.uic.edu/~liub/WebContentMining.html

iv. http://cyberartsweb.org/cpace/ht/lanman/wsm1.htm

v. http://ijcsit.com/docs/Volume%206/vol6issue04/ijcsit2015060404.pdf

****

Prof. Halvadi Homera Nurmamad
Assistant Professor
Svet College
Jamnagar