A Review Paper on Web Mining

Neha Soni

CSE Dept, RIMT, MandiGobindgarh,Punjab , India

ruheesoni@yahoo.co.in

Abstract— **As growing need of World Wide Web in Corporate business, Web mining plays an important role to understand Visitors usage habits. In order to design web site one should first analysis web visitor pattern. In this paper I am presenting three Web mining. Web Usage mining, web Content mining, Web Structure mining.**

Key words - **Web Mining; Web Content Mining; Web Structure Mining; Web Usage Mining, application.**
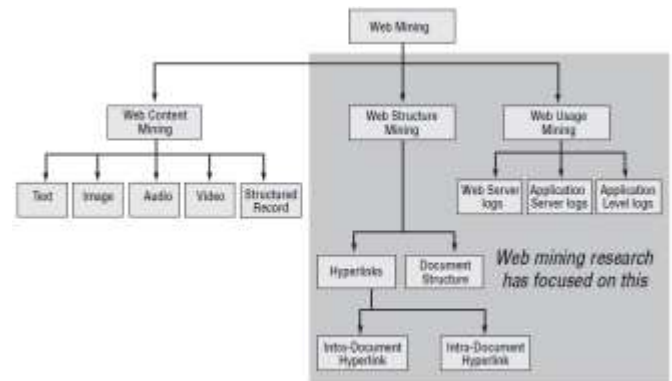
I Introduction

Internet is network of networks. Today many pictures, video are available via internet and with internet one can ease and speed up E commerce business. As growing need of online business, corporate business needs to track user browsing habits to know the interest of users. This paper describes how the web mining technologies make possible for vendor to find the interest of individual customer so to personalize their product.

The next session of paper present an overview of Web mining and its three categories: web usage mining, Web Content mining, Web Structure mining.

II Web Data Mining

Overview

Web Data Mining : Data mining is the technique used by business administrator to get the desired data from data ware house , Data ware house is a collection of raw database which contain information of customer daily needs and usage .Web data mining is also one of the technique of data mining so to extract the web data .Web data contain the information of user browsing pattern which is collected by knowing the page request done by user which can gather by many data sources like knowing user IP address , requested session time/date of users. There are three kinds of web mining categories:

III Web Content Mining

Web Content mining is the process to get information from the content of web documents. Content data is the collection of contents of web pages which contain text, video, audio, pictures lists and tables. Web content mining mostly deal with semi structure data.
Data sources of web content mining are:

Web Page: Web page contain a collection of hyper pages which consist of many content containing advertisement, notices etc.

Search page: Web browser is the medium which is used for search web page. Basically search page is used for searching. Various Search engine used like Google, msn for user search queries.

Result Page: Result page consist of result of web pages. Search engine contain various navigation links of server .Result page is the web page queried by user.

IV Web Structure Mining

Web structure mining is the process of getting data about the structure information of web The Structure information consists of hyperlink, web pages. There are of two types:

Hyperlink: Hyperlink consists of page links which is used by user to navigate. Hyperlink links different pages in same web page or sometime in different web page.

Document Structure: Document Structure is consisting of tree structure type format through which page can link with each other.

The Content of Web Structure Mining is represented as following:

A Link Structure Mining: One of the important and oldest way of searching in Web Mining in which pages are linked as link type classification and Cluster analysis.

B Internal Structure Mining: Internal Structure Mining represents website internal designing and its relationship between another website through model link structure.

C URL Mining: URL Stands for Uniform Resource Locator which is used as pointer for client and links various web pages with each other from same and different location.

V Web Usage Mining

Web Usage Mining is one of the processes of Data mining which is used to extract usage pattern of browsing user. Web Usage mining mainly used by website administrator and business administrator for making better strategies by knowing user habits so one can place advertisement about product to particular user group. Data gather by web Usage Mining is Semi Structured so it need lots of preprocessing to extract the required information. There are data sources for Web Usage Mining which defines as following:

A Web Server Data: Web Server used user log to collect user related data including IP address, page requested pattern, request date and time etc. Log is a file which is used to save user related data in same and different location called as access Log, Referrer Log. These types of Log can only authorize to Web master or administrator person to examine traffic pattern.

B Application Server data: Application server is a log which contains commercial related data about webby using these types of commercial information one can easily enable Ecommerce application according to requirement of user.

C Proxy Server log: Proxy server log is used as data source to extract the behavior pattern of user. Proxy server is act as server which contains multiple time access user data in temporary file called as cookies so to reduce the load of main server by providing same data to user from cookies.

Web Usage mining is represented or categorized in three phases as describe following:

Preprocessing: Preprocessing is the first phase used by web usage mining to extract useful information from semi structure format .It consist of three types:

Usage preprocessing: Data available in this phase is not complete. In this preprocessing data is related with user identification which gets it by its IP address, request time/date.

Content Preprocessing: In this preprocessing content is representing by viewing of pages.

Structure Preprocessing: In this preprocessing pages are link with each other that represent structure of websites.

Pattern Discovery: Pattern Discovery is a second phase which represents access log data in the form of tables, lists, charts and graphs etc.

Pattern Analysis: Pattern analysis is the final phase through which one can get valid information which is required by users.

VI Web Usage Technique

A Association rule: After preprocessing phase, association rule can be used for extract useful information from correlated referenced pages which is referred by one session.

B Clustering: Clustering means same as class which having collection of data/object having same properties or characteristics. Similarities between different objects can be collected by online survey. Clustering of users help business administrator to plan their strategies according to user interest by knowing user navigation habits by only one group.

C Classification: Classification is same as clustering but it is predefined class. Many webmaster use this type of technique to describe one particular class for user such as class having user interest in classical music.

VII Application of Web Mining

A E Commerce: E commerce means online business, by using web mining technique easily understand user behavior pattern and placed advertisement.

B Information Filtering: One can used web mining for filtering information according to user needs.

C Fraud Detection: Web mining contains data of user including IP address, its signature. so one can easily identified fraud by checking their details.

D Website Designing: Web mining also helps website administrator to design a website by knowing user browsing pattern.

E Search Engine: There are various search engines like Google, msn etc. Web mining also helps by providing user related access pages data.

VIII Conclusion

This paper provides the survey on web mining, its three categories and application. In this paper a general overview about usage of web mining is describe how its helps Web master and Business administrator to plan its strategies, the key topics are: Web Content mining, Web structure mining, Web usage mining, Web mining technique: Association rules, Clustering, Classification .

References

[1] Manoj pandia" A Review Of Trends In Research On Web Mining"

[2] Mr. Dushyant B. Rathod " A Review on Emerging Trends of Web Mining and It's Applications | ISSN: 2321-9939 "

[3]Web mining and knowledge discovery of usage pattern..

[4]Jagdeep srivastava"Web mining-concepts , application,and research directions

[5]"A survey paper on web usage mining and pattern discovery"

[6] Anand Sharma"Web usage mining:data preprocessing, pattern discovery and pattern analysis on the rti web data".

[7]Priyanka bharat"A review paper on web usage mining and future predtiction"

[8] K. R. Suneetha " Identifying User Behavior by Analyzing Web Server Access Log File "