IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Data and Web Mining

B B DEEKSHIKA
dept. Computer science and engineering.
RNS INSTITUTE OF TECHNOLOGY
Bengaluru(Karnataka), INDIA

CHANDANA D
dept. Computer science and engineering.
RNS INSTITUTE OF TECHNOLOGY
Bengaluru(Karnataka), INDIA

CHAITHRA BK
dept. Computer science and engineering.
RNS INSTITUTE OF TECHNOLOGY
Bengaluru(Karnataka), INDIA

CHAITHRASHREE H
dept. Computer science and engineering.
RNS INSTITUTE OF TECHNOLOGY
Bengaluru(Karnataka), INDIA

ANUSHA PY
dept. Computer science and engineering.
RNS INSTITUTE OF TECHNOLOGY
Bengaluru(Karnataka), INDIA

Abstract

Data mining and web mining are two fields that are likely to be related that involve extracting useful information from large datasets. Data mining refers to the process of analysing data from different perspectives and discovering patterns, relationships, and trends. This process is often used in business intelligence, market research, and scientific discovery. Web mining, on the other hand, focuses on extracting information from the World Wide Web(WWW). It includes analysing web content, link structures, and usage data to discover patterns and trends. Web mining is used in a variety of applications, such as search engine optimization, e-commerce, and social network analysis. In order to analyse internet usage, data is gathered from the web browser log file using data mining methods. Nowadays, there is a huge amount of information that is constantly rising quickly. Data miners use these methods to find and retrieve knowledge from online data.

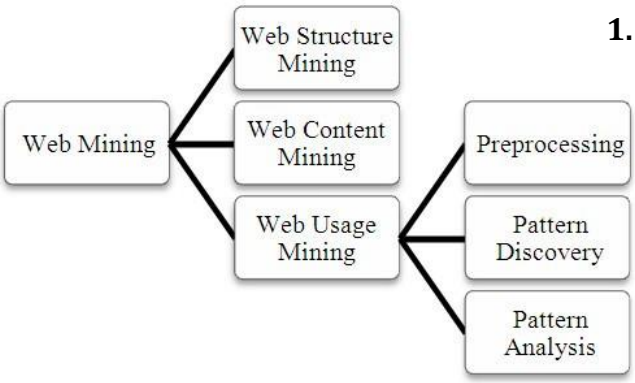
Keywords: Data Mining, Web Usage Mining, Web Mining.

1 Introdution

Web mining is a methodology that uses algorithms for data mining to easily learn and obtain information from internet papers as well as services.Finding useful patterns and information from web pages is a process known as web mining.They are a data mining subfield that focuses on examining web data to extract useful knowledge and understanding. Search engine optimization, e-commerce, social network analysis, and online personalization are just a few of the various benefits for web mining.It[1] is the combination of data collected by earlier data mining mechanisms and procedures with data collected through the Internet (WWW).Web Usage Mining can be used to help understand behavior of customers. Web mining examines web data using data mining, artificial intelligence, chart technology, and other methods to better represent customers. It also tracks users' browsing habits and extracts their usage patterns[2].

fig.1: Framework of Web Mining

1.1



The above Fig 1 shows the Framework of Web mining.

Web mining is currently one of the most significant subfields in this area of computer and information sciences. web mining is to extract patterns from information [3] [4].Web mining models are used to detect various types of web data, including text, image, and multimedia data. Web content mining, which usually includes retrieval of necessary data from web sites. web structure analysis, which involves looking at how websites are linked together. web usage mining, which includes the examination of user behaviour on the web, are examples of these techniques.Web usage mining works with secondary information produced from user communication, in comparison to web content and structure mining, that also makes use of the web's primary data. Web usage mining searches for database transactions on a website, weblogs, click streams, and user interactions with a web server[5].Data mining is just a WUM software that retrieves information from blog server web pages. The process of comparing information from various sources and reducing it to relevant information is known as data mining. It is an interdisciplinary field that incorporates methods and techniques from a large range of fields which includes information extraction, and natural language processing.

A technique for finding patterns in online interconnections is called web structure mining.

Types of Web Mining

Web content mining, web structure mining, and web utilization mining are the three subcategories of web mining. Each kind of web mining examines various kinds of online data and can produce a variety of information and facts.

Web content mining :The method of eliminating relevant information from websites is known as web content mining. Text, images, and image collections can all be included. Natural language processing, text mining, and image and video analysis are some of the methods used in it.The insights gained from web content mining can include sentiment analysis, topic modeling, and content categorization.

Web structure mining : It involves analyzing the interconnection structure of web pages. This can include analyzing the links between web pages, the types of links, and the patterns of links. Techniques used in web structure mining include graph theory, link and social network analysis. The insights gained from web structure mining can include identifying important web pages, detecting web communities, and understanding the relationships between web pages.

Web usage mining :It involves analyzing user behavior on the web, such as clickstream data, cookies, and server logs. Techniques used in Clustering, association rule mining, and pattern matching mining are a few examples of online usage mining.The insights gained from web usage mining can include identifying user preferences, predicting user behavior, and improving website usability.

Web mining methods can be used in a variety of applications, including e-commerce, social network analysis, and search engine optimization, to obtain valuable knowledge and trends from various kinds of web data.

2 Web Usage Mining

Web usage mining refers to the method of gathering useful data and understanding from web usage data, which is obtained from user-web application exchanges. It involves the examination of website visitor behavior, such as the pages they view, the links they select, how long they stay on each webpage, and other applicable measurements.Data preprocessing, pattern discovery, and knowledge representation are the three stages that generally make up the web usage mining process. The raw information is collected, cleaned up, and put into a structure that is simple to be evaluated during the data preprocessing stage. The identified patterns are represented visually and in a manner that is simple for users to understand during the knowledge representation stage. It is also known as web log mining.it entails finding and eliminating interesting information from extensive web pages, the Internet. Web servers assemble data about the user's interactions each time a request for a resource is made.

2.1 Web Usage Mining Process

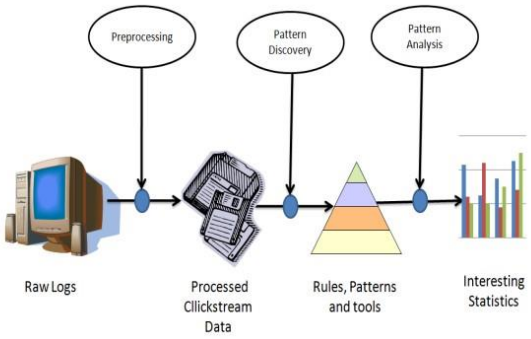


fig.2.1 Web Usage Mining Process

The above Fig.2.1 shows the process of Web Usage mining.

The three techniques used in online usage mining are pattern discovery, pattern analysis, and web data preprocessing.

i)Preprocessing : Data preparation involves the task of transforming raw data into the format needed by the data mining method. To meet user expectations, data must be preprocessed into a format that can be processed more quickly and simply. This phase typically marks the beginning of the data mining process. In internet use mining, pre-processing is done to increase the value of the information. This can be done by eliminating the randomly generated data from the web data collection after extracting and processing the data.

ii)Pattern Discovery : In online usage mining, which combines data mining methods and approaches, pattern finding is the key component. Customers' searches and website visits are necessary during the phase of discovering intriguing trends. iii)Pattern analysis : The most intriguing patterns are selected from the patterns that were recovered during the pattern discovery phase through preprocessing during the pattern analysis phase. Predicted additional data from the same source is a feature of pattern analysis. Text, audio, video, and image data are just a few examples

2.2 Techniques in Web Usage Mining

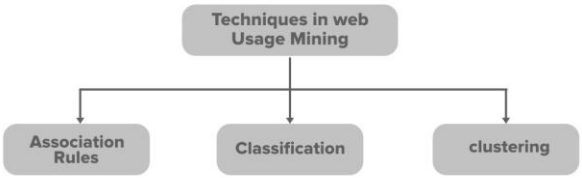


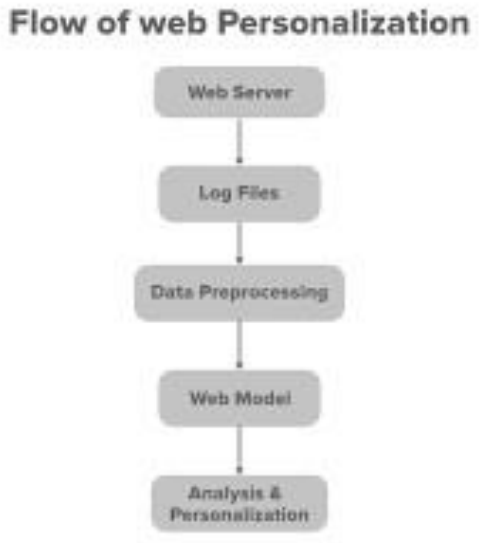
fig.2.2:Techniques in Web Usage Mining

The above fig 2.2 shows the techniques involve in the Web Usage Mining.

i)Association Rules : Association Rules is the method that Web usage mining uses the most. Essentially, this method focuses on the connections between websites that regularly occur in users' sessions. A single server session is always created from all of the pages that were viewed at once. Utilizing access logs and association rules, websites can be rebuilt. In general, access logs detail requests that are approaching the web server. This method's fundamental disadvantage is that it may create too many sets of rules, some of which may turn out to be completely irrelevant. They might not also brought to use in the future.

ii)Classification : Classification can be implemented using a variety of methods, including Support Vector Machines, Decision Trees, etc. For instance, a consumer can be classified into frequent and nonfrequent groups or categories depending on their purchase patterns over the past six months. Other circumstances may also allow for multiclass. Assigning specific data to multiple, established classifications is the core goal of categorization. Online use mining's main objective is to develop user or client profiles that are connected to a certain class or category. Extraction of the better qualities that will work best for the appropriate class is required in order to reach this precise goal.

iii)Clustering: Clustering is a method for assembling a collection of entities with related characteristics. There are basically 2 different kinds of clusters: the usage cluster and the page cluster. Based on the usage data, it is simple to cluster the pages. User clustering frequently creates groups of people with comparable internet activities. The fundamental idea behind page clustering is to easily access information across web pages.



2.3 Web Usage Mining Applications

i)Personalization of Web Content : The World Wide Web is home to a vast and expanding body of knowledge. The fundamental problem is that, even though people's specialized demands are expanding daily, they frequently don't get the search results they want. Thus, web customization offers a remedy for this.

The below fig 2.3 shows the flow of web personalization.

fig.2.3:Flow of web Personalization

The best way to describe web customization is to say that it caters to the user's demands following their interests and navigational behavior. Checkbox personalization, recommender systems, etc. are all examples of web personalization. Due to their widespread use, recommender systems are used by many companies.

ii)E-commerce : Web-based enterprises' use of mining to analyse web activity is significant. Since gaining new customers,

keeping existing ones, upselling, etc. are their primary priorities. Web-based firms that want to build long-lasting client connections must use web usage mining, which offers a lot of data on customer interests. It also offers suggestions on how the company might enhance specific aspects of its website design.

iii)Prefetching and Catching : Prefetching, as the name implies, is the process of loading data ahead of time to minimize the time needed for it. Prefetching and caching strategies can be created using the data gathered through web usage mining to dramatically speed up server response times.

3 Web Structure Mining

Web structure mining is the process of accomplishing the organization of links between web pages in search of useful information and trends. In order to find connections, dependencies, and groupings among web pages and links, it involves studying their organisation, topology, and connectivity.

The process of developing structural summaries about websites and web pages by analysing the linkages is known as web structure mining. These structural abstracts can be used to enhance web-building activities.

Web structure mining has a number of advantages, such as raising the standard of search results, locating web sites and communities, identifying spam and fraud, and enhancing website design. However, it also creates privacy issues because the information gathered may contain sensitive data about websites and the people who own them. As a result, it's significant to stick to moral and legal standards when gathering and examining online structure data.

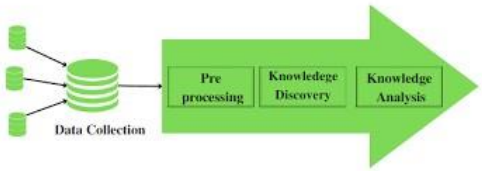


fig.3:Web Structure Mining Process

The above fig 3 shows Web Structure Mining phases (or) process.

The web structure mining process consists of four phases.

Data Collection.

Preprocessing.

Knowledge Discovery.

Knowledge analysis.

i)Data collection : Any mining technique's first step is to gather the data needed for analysis. Data collection in web structure mining refers to gathering links from online sites connected to seed URLs on different servers.

ii)preprocessing : implements a set of methods for cleaning up data in web links files, validating links, identifying unique connections, and finishing links. iii)Knowledge Discovery : analyzing data using various data mining techniques, including clustering, pattern analysis, association, analytical explanation, and related ones.

iv)Knowledge analysis : After confirming information discovery from online links, remove irrelevant data and estimate and consider the user-interesting pattern.

4 Data Mining

As information technology has advanced, a massive amount of databases are now being created. Data mining is a technique for removing information from numerous datasets. It is commonly referred to as the knowledge discovery process, data analysis, knowledge extraction, or knowledge mining from data.

The technique of examining huge amounts of information to find patterns and trends is known as data mining. Effective data collection and computer processing are necessary for data mining.It is used in a wide range of situations, such as database marketing, preventing scams, and managing credit risks.

A big collection of data is transformed into usable output through the application of various data mining techniques. It is the rational approach. This approach's primary goal is to look for a previously unidentified pattern.

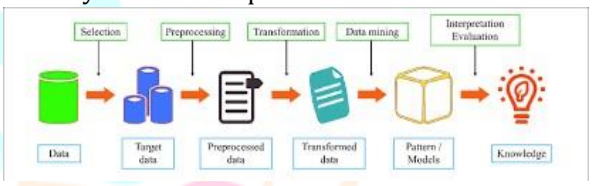


fig.4:Steps for Data Mining The above fig 4 shows the steps for Data Mining.

This process comprises three steps:

1.Exploration

2.Deployment 3. Pattern recognition.

5 Conclusion

One of the most common methods for promoting is the website. The attitudes of online users are gathered through the mining of log files used for web usage. Security, identity verification, and the detection of unusual access to protected data are all possible uses for the method known as web usage mining. Web Usage Mining can be used to better a search's quality and the effectiveness of upcoming accesses by identifying the required connections by identifying users' frequent access patterns.

6 Reference

[1]Hendrik Blockeel and Raymond Kosala published "Web mining research: A survey." 2, no. 1 of ACM Sigkdd Explorations Newsletter (2000): 1–15.

[2]Chen, M. S., Park, 1. S., and Yu, P. S., "Efficient Data Mining for Path Traversal Patterns", IEEE Transactions on Knowledge and Data Engineering, MarchiApril, 1998, pp. 209-221.

[3]"The research of Web mining," by Liu, Lizhen, Junjie Chen, and Hantao Song. published in 2002's Intelligent Control and Automation. 4th World Congress on Proceedings, volume 3, pages 2333–2337. IEEE, 2002.

[4]Web mining in soft computing framework: relevance, state of the art, and future directions, by Sankar K. Pal, Varun Talwar, and Pabitra Mitra. 11,63–1177 in IEEE Transactions on Neural Networks 13, no. 5 (2002).

[5]"Innovative Pre-Processing Technique and Efficient Unique User Identification Algorithm for Web Usage Mining," R. Sriram and R. Malliga. Vol. 6, Issue. 02, pp 85–91, International Journal of Advanced Research in Computer Science and Software Engineering, 2016.

[6]https://www.geeksforgeeks.org.

[7]https://www.tutorialspoint.com.