

WEB MINING USING CLOUD COMPUTING

¹MR.PARAG D PATEL

¹ Department of Computer Engineering, C. U. Shah Technical Institute of Diploma Studies, C. U. Shah University, Surendranagar, India.

patel_paragkumar@gtu.edu.in

ABSTRACT : In this paper we present the web mining using Cloud Computing Technology. Web mining includes how to extract the useful information from the web and gain knowledge using data mining techniques. Here so many resources and techniques are available i.e. web content mining, web structure mining, web usage mining and access through the web servers. Web mining techniques and applications are much needed in cloud computing. The implementation of these techniques through cloud computing will allow users to retrieve relevant and meaningful data from virtually integrated data warehouse which reduces cost and infrastructure.

KEY WORDS : Data Mining, Web Mining, Data Warehouse, Knowledge Discovery, Cloud Mining, Web Content Mining, Web Structure Mining, Web Usage Mining.

1. INTRODUCTION

Web mining - is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining. Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet using cloud computing . Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered. Several data mining methods are used to discover the hidden information in the Web. However, Web mining does not only mean applying data mining techniques to the data stored in the Web. The algorithms have to be modified such that they better suit the demands of the Web [1].

2. HISTORY OF WEB MINING

Web mining techniques are the result of long process of research and product development. This evolution began when business data was first stored on computers & internet, continued with improvements in data access and more recently real-time technologies that users in the WWW (World Wide Web) allow us to navigate through our data.

Data gathered from surveys, or input from several independent or networked locations via Computer & tapes are define Data collection. Use the data storage typically refers to software and related activities, the retrieving, or stored in a database or other data source acting on.

In the evolution from business data to business information each new step has built upon the previous one. For example, the ability to store large databases is critical to web mining. From the user point of view, the five step listed in Table I were revolutionary because they allowed new business question to be answered accurately and quickly.

Basically data mining technique are used in web mining. Web mining is extended version of data mining. Data mining is work upon Off-Line whereas Web mining is work upon On-Line. In data mining data stored in (database) data warehouse and in web mining data stored in server database & web log. The main component of Web Mining Technology have been under development for decades, in research area such as internet, artificial intelligence, and machine learning.

3. APPROCH OF WEB USAGE MINING

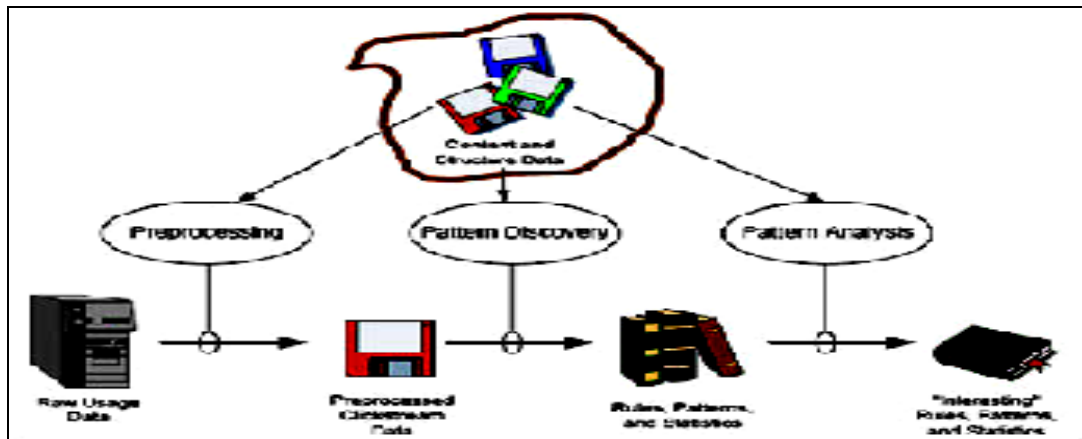


Figure 1: Concept of Web Usage Mining

The web usage mining generally includes the following several steps: data collection, data pretreatment, and knowledge discovery and Pattern analysis.

Data collection:

Web usage mining is the process of extracting useful information from server logs e.g. use Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered: Web Server Data: The user logs are collected by the Web server. Typical data includes IP address, page reference and access time. Application Server Data: Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs[3]. Application Level Data: New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above in the figure() [3].

Data preprocessing:

Web Usage Mining in cloud computing is one of the categories of data mining technique that identifies usage patterns of the web data, so as to perceive and better serve the requirements of the web applications.

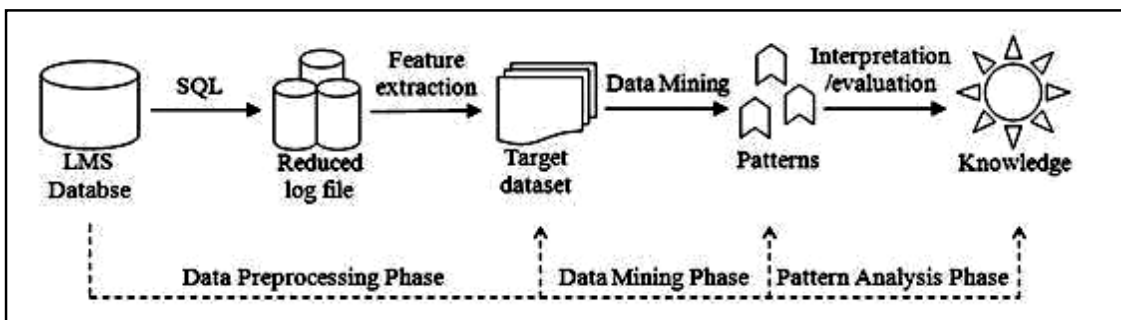


Figure 2: Data Preparation

The working of WUM involves three steps - preprocessing, pattern discovery and analysis. The first step in WUM - Preprocessing of data is an essential activity which will help to improve the quality of the data and successively the mining results. This research paper studies and presents several data preparation techniques of access stream even before the mining process can be started and these are used to improve the performance of the data preprocessing to identify the unique sessions and unique users in cloud computing . The methods

proposed will help to discover meaningful pattern and relationships from the access stream of the user and these are proved to be valid and useful by various research tests. The paper is concluded by proposing the future research directions in this space [2]. In the Data Pretreatment work mainly it includes data cleaning, user Identification and path completion.

Data Cleaning

The most important task of the Web Usage Mining in cloud computing process is data preparation. This process is diagrammatically represented in Fig(2) . The success of the project is highly correlated to how well the data preparation task is executed. It is of utmost importance to ensure, every nuance of this task is taken care of. This process deals with logging of the data; performing accuracy check; putting the data together from disparate sources; transforming the data into a session file; and finally structuring the data as per the input requirements. The data used for this project is from the RIT Apache server logs, which is in the Common Log File format. This access log includes the agent and the referrer in the data as one of the attributes [4].

Path completion

An implementation of data preprocessing system for Web usage mining and the details of algorithm for path completion are presented. After user session identification, the missing pages in user access paths are appended by using the referrer-based method which is an effective solution to the problems introduced by using proxy servers and local caching. The reference length of pages in complete path is modified by considering the average reference length of auxiliary pages which is estimated in advance through the maximal forward references and the reference length algorithms. As verified by practical Web access log, the proposed path completion algorithm efficiently appends the lost information and improves the reliability of access data for further Web usage mining calculations [5].

Knowledge Discovery

In general, knowledge discovery can be defined as the process of identifying interesting new patterns in data. These patterns can be, e.g., relations, events or trends, and they can reveal both regularities and exceptions [3].

4. LITERATURE SURVEY

I read many article to web usage mining these are following: **A Framework for Personal Web Usage Mining:** In this paper, I got to mine Web usage data on client side, or personal Web usage mining, as a complement to the server side Web usage mining. By mining client side Web usage data, more complete knowledge about Web usage can be obtained [2].

A Research Area in Web Mining: This paper also discusses an application of WUM, an online Recommender system that dynamically generates links to pages that have not yet been visited by a user and might be of his potential interest. Differently from the recommender systems proposed so far, it does not make use of any off-line component, and is able to manage Web sites made up of pages dynamically generated [3].

Cloud Mining: This paper also discussed about Web usage mining and user behavior analysis using fuzzy C-means clustering: In this paper I got methodologies used for classifying the user using Web Usage data. This model analysis the users behaviors and depend on the interests of similar patterns provides appropriate recommendations for active user [1].

Discovery and Applications of Usage Patterns from Web Data: This paper provides an up-to-date survey of the rapidly growing area of Web Usage mining. With the growth of Web-based applications, specifically electronic commerce, there is significant interest in analyzing Web usage data to better understand Web usage, and apply the knowledge to better serve users [4].

Web Mining Using Cloud Computing: This paper provide present the technology of cloud computing using web mining .Web mining include how to extract the useful information from the web and gain knowledge using data mining techniques. Here so many online resources are available i.e. web content mining and access through the web servers [5].

5. RELETED WORK

Many researchers have looked for way of represent the web mining and future of web mining in Cloud Computing. Some of these are said that cloud mining is the future of web mining. This paper describes the web usage mining in Cloud Computing technology. Web usage mining model is a kind of mining to server logs. Web Usage Mining plays an important role in realizing enhancing the usability of the website design, the improvement of customers' relations and improving the requirement of system performance and so on. Web usage mining provides the support for the web site design, providing personalization server and other business making decision, etc[7]. Web usage mining in Cloud Computing is clearly one of today's most seductive technology area in research field due to its cost efficiency and flexibility. However, despite increased activity and interest, there are significant, persistent concerns about cloud computing that are impeding momentum and will eventually compromise the vision of cloud computing as a new IT procurement model SaaS (Software-as-a-Service) is used for reducing the cost of web mining and try to provide security that become with cloud mining technique. Now a day we are read y to modify the framework of web mining for demand cloud computing. In terms of —mining clouds, the Hadoop and Map Reduce communities who have developed a powerful

framework for doing predictive analytics against complex distributed information sources[9]. The logs are preprocessed to group requests from the same user into sessions. A session contains the requests from a single visit of a user to the Web site.

6. ONLINE WEB USAGE MINING IN CLOUD SYSTEM

Web based recommender systems are very helpful in directing the users to the target pages in particular web sites. Moreover, Web usage mining cloud model systems have been proposed to predict user's intention and their navigation behaviors. In the

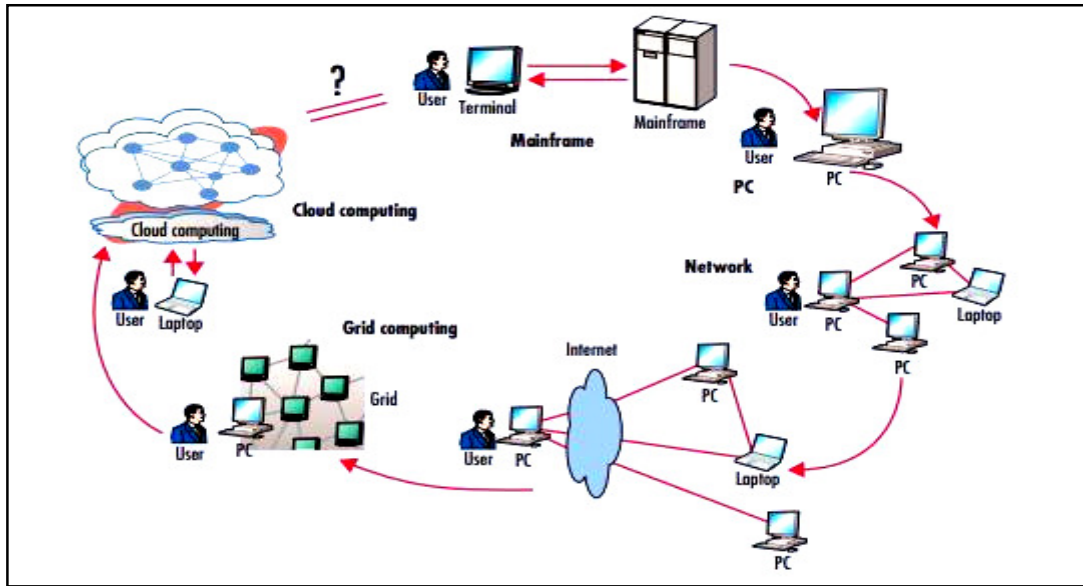


Figure: 3: Cloud Computing Aspect

Following, we review some of the most significant WUM systems and architecture that can be compared with our system. Cloud system proposed a Cloud model for navigation pattern mining through Web usage mining to predict user future movements. The approach is based on the graph partitioning clustering algorithm to model user navigation patterns for the navigation patterns mining phase [8].

7. CONCLUSION

Cloud Computing is a broad term that describes a broad range of services. As with other many vendors have seized the term “Cloud” and are using it for products that sit outside of the common definition. In order understand how the Cloud can be of value to an organization, it is first important different components. Since the Cloud is a broad collection of services, organizations can choose where, when, and how they u Cloud Computing. In this paper I explained the different types of Cloud Computing Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) with the help of web using mining in cloud Services [10].

8. REFERENCES

1. Web mining, “www.wikipedia.com”, 10-02-2015.
- 2.”A Framework for Personal Web Usage Mining”.
3. Rajni Pamnani, Pramila Chawan Department of computer technology, VJTI University, Mumbai “Web Usage Mining: A Research Area in Web Mining”.
4. Kaikala Anjani Sravanthi1, “Web Mining Using Cloud Computing” , ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.
5. Mobasher, H. Dai, T. Luo, and M. Nakagawa. Discovery and evaluation of aggregate usage profiles for web personalization. *Data Mining and Knowledge Discovery*, 6(1):61 {82, January 2002.
6. Jaideep Srivastava et al , “Discovery and Applications of Usage Patterns from Web Data” , Jan 2000 Volume 1, Issue 2 -page 12.
7. Rajni Pamnani, Pramila Chawan Department of computer technology, VJTI University, Mumbai “Web Usage Mining: A Research Area in Web Mining”.
8. Kaikala Anjani Sravanthi1, “Web Mining Using Cloud Computing” , ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013.