

CLARIFICATION OF WEBBING THROUGH CLOUD COMPUTING TECHNOLOGIES

Palakollu. Divya

Department of Computer Science and Engineering

Asst Professor at Samskruti College of Engineering and Technology

Cloud Computing

ABSTRACT

In this paper we present the web mining using Cloud Computing Technology. Web mining include how to extort the functional in order from the web and expand knowledge using data mining technique. So many assets and technique are existing i.e. web content mining, web usage mining and right to use through the web servers. Web mining techniques and application are much crucial in cloud computing. The implementation of these techniques through cloud computing will allow users to recover appropriate and consequential data from virtually incorporated data warehouse which reduce the cost and infrastructure.

INTRODUCTION

Web Mining is type of data mining which is used for web. Web Mining has become an emerging and an important trend because of the main reason that today storage of data has become enormous that retrieving and processing them has become an overhead. Two different approaches were taken in initially for defining Web mining they were 'process-centric view' and 'data-centric view'. Process centric accounted the sequence of tasks whereas data centric accounted for the types of web data that was being used in the mining process. The second approach is taken into consideration widely in recent times. Web Mining has been categorized into three major distinct categories: Web Content Mining, Web Usage Mining and Web Structure Mining. The implementation of these categorises on World

Wide Web have been well reviewed. But usage of these on the CC is newly clubbed technology

ABOUT WEB DATA MINING

Web data mining can be classified into three. They are Web content mining, Web structure mining and Web usage mining.

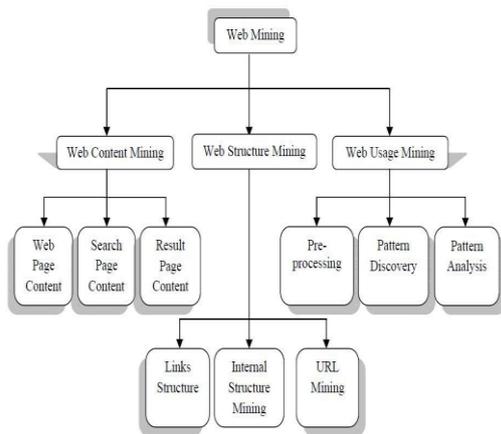
- a) Web Content Mining
- b) Web Structure Mining
- c) Web Usage Mining

i. Web Content Mining

Web Mining is basically extract the information on the web. Which process is happen to access the information on the web. It is web content mining. Many pages are open to access the information on the web. These pages are content of web. Searching the information and open search pages is also content of web. Last accurate result is defined the result pages content mining..

ii. Web Structure Mining

We can define web structure mining in terms of graph. The web pages are representing as nodes and Hyperlinks represent as edges. Basically it's shown the relationship between user & web. The motive of web structure mining is generating structured summaries about information on web pages/webs. It is shown the link one web page to another web pag



Pattern Discovery- According the data pre-processing discovered the knowledge and implements the techniques to discover the knowledge like as machine learning and data mining procedures are carried out at this stage.

Pattern Analysis- pattern analysis is the process after pattern discovery. Its check the pattern is correct on the web and how to implement on web to extract the information on your web search / extract knowledge from the web.

(Classification of Web Mining)

iii. Web Usage Mining

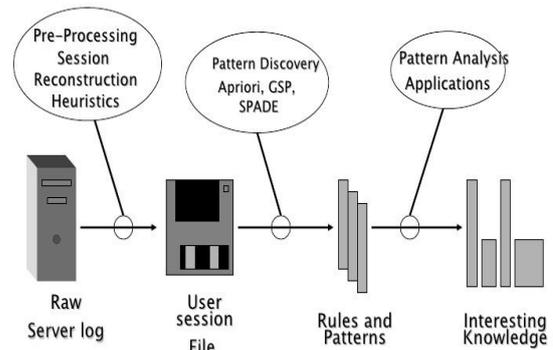
It is discovery of meaningful pattern from data generated by client server transaction on one or more web localities. A web is a collection of inter related files on one or more web servers. It is automatically generated the data stored in server access logs, refers logs, agent logs, client sides cookies, user profile, meta data, page attribute, page content & site structure. Web mining usage aims at utilize data mining techniques to discover the usage patterns from web based application It is technique to predict user behaviour when it is interact with the web. Web usage mining is categories in three phases:-

- Pre-processing
- Pattern Discovery
- Pattern Analysis

Pre-processing- According to client, server and proxy server it is first approach to retrieves the raw data from web resources and processed the data .it is automatically transformed the original raw data.

Web Mining

Phases of Web Usage Mining



RELATED WORK

Numerous researchers have looked for mode of signify the web mining and prospect of web mining in Cloud Computing. A few of these are alleged that cloud mining is the upcoming 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 enhancement of customer associations and recuperating the requirement of system performance and so on. Web usage mining provides sustain for the web site design. Web usage mining in Cloud Computing is visibly one of today's most seductive technology areas in investigate field due to its cost effectiveness and suppleness. However, regardless of amplified interest and activity, there are persistent, significant concerns about cloud computing that are impede momentum and will in due course compromise the visualization of cloud computing as a new IT procurement model.

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.

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.

B] 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. 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 .In the data pretreatment work, mainly include data cleaning, user identification, session identification and path completion

1)Data Cleaning:

The most important task of the Web Usage Mining in cloud computing process is data preparation. 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

2) 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

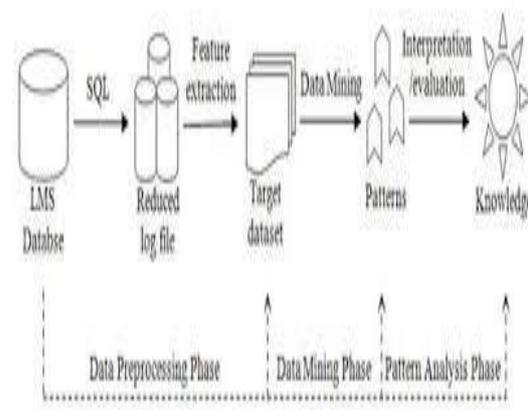
servers and local caching. The reference length of pages in complete path is modified by considering the average reference length of auxiliary pages which is estimated in advance through the maximal forward references and the reference length algorithms. As verified by practical Web access log, the proposed path completion algorithm efficiently appends the lost information and improves the reliability of access data for further Web usage mining calculations

3) 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

D) Pattern analysis:

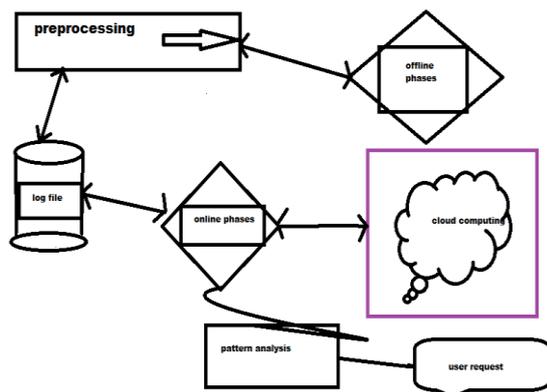
Challenges of Pattern Analysis are to filter uninteresting information and to visualize and interpret the interesting patterns to the user. First delete the less significance rules or models from the interested model storehouse; Next use technology of OLAP and so on to carry on the comprehensive mining and analysis; Once more, let discovered data or knowledge be visible; Finally, provide the characteristic service to the electronic commerce website



WEB USAGE MINING

ONLINE WEB BASED MINING IN CLOUDSYSTEM

Web based recommender system is extremely accommodating in directing the users to aim pages in particular web sites. Moreover, Web usage mining cloud model systems have been proposed to predict user's intention and their navigation behaviours. In the 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.



CONCLUSION

As the Web has become a major source of information, techniques and methodologies to extract quality information is of paramount importance for many Web applications and users. In this paper, various concepts are outlined to extract the useful information from the web. Web

design patterns are useful tools for web data mining. Web pages are analysed to find out which useful information is included in web data. Web Data mining technologies provided through Cloud computing is an absolutely necessary characteristic for today's businesses to make proactive, knowledge driven decisions, as it helps them have future trends and behaviours predicted. This paper provides an overview of the necessity and utility of web data mining in cloud computing. As the need for data mining tools is growing every day, the ability of integrating them in cloud computing becomes more and more stringent. The effective management of information kept over cloud for knowledge and business management could be achieved if the web mining techniques over virtually integrated warehouse are efficiently practiced by an end users. The implementation of the web mining techniques through cloud computing will allow users to retrieve relevant and meaningful data from virtually integrated data warehouse which reduces costs of infrastructure and storage. The Future work of web mining is to introduce a hierarchy on the information about the website. Now we also work for the process mining and try to combine usage mining with structure mining. We can also go for the mining from cloud. Whenever we work on mining over cloud computing that time we hesitate for the cost but that comes very less by cloud mining. So, we can say that cloud mining can be seen as future of web mining

FUTURE WORKS

There are so many data mining algorithm for discovering, sharing and utilising the knowledge existing on the web , but there are also quite a few existing problems that should be taken into consideration, such as the inadequate

utilization of network resources and the lack of individualisation of the existed platforms. So, the future work that has to be conducted is how to improve the effectiveness of data mining methods; dynamic data and knowledge of the data mining; network and distributed environment, such as data mining; the development to adapt more types of data to allow for the noise of the dig methods. As the development of World Wide Web and its usage grows, it will continue to generate ever more content, structure, and usage data and the value of Web mining will keep increasing. Research needs to be done in developing the right set of Web metrics, and their measurement procedures, extracting process models from usage data, understanding how different parts of the process model impact various Web metrics of interest, how the process models change in response to various changes that are made for changing stimuli to the user, developing Web mining techniques to improve various other aspects of Web services, techniques to recognize known frauds and intrusion detection and various web mining algorithms for supporting Big Data on the web and user friendly services. Future extensions using cloud computing will include adding up of more data mining services to be provided by the cloud server.

REFERENCES

- [1] M. Rathamani et al, "Web usage mining and user behaviour analysis using fuzzy C-means Clustering",
- [2] Data pre-processing;
- [3] Web mining , "www.wikipedia.com"
- [4] Jaideep Srivastava et al , "Discovery and Applications of Usage Patterns from Web Data" , Jan 2000 Volum
- [5] Abdelhakim Herrouz et al. "Overview of Web Content Mining Tools" The International Journal of Engineering and Science (IJES) (2013) Volume: 2 Issue: 6 and Refer-ences there in.
- [6] Miguel Gomes da Costa Júnior and Zhiguo Gong. "Web Structure Mining: An Introduction" Proceedings of IEEE International Conference on Information Acquisition (2005)
- [7] ORACLE, —Oracle Data Mining Mining Techniques and Algorithms, Link:<http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-techniques-algorithms-097163.html>.
- [8] Faten Khalil, Jiuyong Li and Hua Wang —A Framework of Combining Markov Model with Association Rules for Predicting Web Page Accessesl ,Proc. Fifth Australasian Data Mining Conference (AusDM2006), CRPIT Volume 61,177-184.
- [9] Wu, K.L. Yu, P. S. Ballman, A. —A Web usage mining and analysis tool—, IBM Systems Journal, 2010
- [10] Michael Jennings, What are the major comparisons or differences between Web mining and data mining? Information Management Online, June 25, 2002.
- [11] Deyi Li, Kaichang Di, Deren Li and Xuemei Shi, —Mining association rules with linguistic cloud modelsl, Research and Development in Knowledge Discovery and Data Mining ,Lecture Notes in Computer Science, 1998.
- [12] Ajay Ohri —Data mining through Cloud Computingl. <http://knol.google.com/k/data-mining-through-cloud-computing#> See on Dec. 2010.
- [13] Etzioni, O. —The World Wide Web: Quagmire or gold minel, Communication of the ACM, Vol. 39, No. 11, pp. 65-68, 1996.
- [14] Sravan Kumar, D. and Naveena Devi, B. —Learner's Centric Approach for Web Miningl et

al. (IJCSIT) International Journal of Computer
Science and Information Technologies, Vol. 1(