International Journal of Research in Advent Technology, Vol.7, No.1, January 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

A Performance Analysis of Web Usage Mining Tools

Varun Malik[1], Dr. Sanjay Singla[2]
Phd Scholar, Professor
Panjab Technical University, Ggs College of Modern Technology
varunmalik.mrce@gmail.com , ss.jnujaipur@gmail.com

Abstract: Web Usage Mining is a prominent name in today's world of the Internet and is used to discover interesting patterns from web server log data. Web usage mining is a branch of Web Mining which analyzes user behavior across the internet. In this paper, we summarize the access patterns of the three major tools of web usage mining, namely, WEBMINER, WEBKIV and WEKA and identify which tool has more features and compatible with today's user access data. We compare all web usage mining techniques across these tools and determine which tool gives better result.

Keywords-Web Mining; Web Usage Mining; WEBMINER; WEBKIV, WEKA;

1. INTRODUCTION

In today's era of technology, the WWW has played an important role as a resource of information since its creation in 1989. The WWW is a vast repository of data and information. Over this data, when we applied techniques to filter out desirable information, as data mining techniques are applied over web data, it is termed as Web Mining. Web Mining is a very important technique of data mining through which we filter out web data[8]. Every page is accessed on the web is continuous processing of information as per users click. Web usage mining, has motivated the discovery of interesting patterns and new rules to identify of hidden pages.

Web usage mining is part of web mining in which we uncover compulsive patterns from web server. In web mining, firstly we apply preprocessing of data followed by a transformation phase to discover new patterns or to analyze web data by applying data mining techniques. For web usage mining, there are a number of tools and techniques available to predict the behavior of users in terms of how they surf the Internet. Web usage mining keeps extracting statistical data or information from web servers[2]. In today's world, most organizations are entirely dependent on the Internet for their business, so that the importance of Web usage mining is increasing day by day. Web Usage Mining is the approach in Web Mining through which we determine user statistics and other resources by which users discover new patterns for their work and business.

2. WEB USAGE MINING

Web usage mining is the branch of Web mining which is used for summarizing meaningful user access patterns' data from server log files. Log files are stored on the server and they keep records of data of every user who visits the web pages. New patterns of data can be discovered for E-commerce websites, and product-oriented and events on websites. In Web usage mining data is processed from Web server log files, where access permission and privacy of data are major issues. Web usage mining can be categorized into mainly three phases of processing.

- Pre-Processing
- Pattern Discovery
- Pattern Analysis

These phases are described next.

2.1 Pre-Processing

Preprocessing is first step in web usage mining where data is pre-processed by applying techniques such as integration, cleaning the data, filtering the data and transforming the data, in such a manner that users eliminate irrelevant data from log files. Preprocessing is a time-consuming step due to the vast variety of server log files. That is why, web usage mining takes more time to process when compared with other web mining techniques. In this phase of preprocessing, data cleaned via user identification and session identification. A session is described as group of activities performed by a user when user navigating through a given website. It is a complex task to identify the session from the raw data, because web server logs files not always contained all the information required. Web server logs files do not contained desired information for reconstructing user sessions. for example, time-oriented heuristics can be used after analyzing the sessions.[4]

2.2 Pattern Discovery

This is the second phase of web usage mining following Preprocessing of web server log data. Pattern discovery act as most useful and important process in web usage mining through we mined the usefulness of the data. This process is most important techniques under association rule. Data mining methods are more reliable and efficient for the discovery of hidden patterns[6]. But today mostly research is in the direction of supervised methods while unsupervised methods are not being applied for pattern extraction from server log files.

2.3 Pattern Analysis

Pattern Analysis used in web usage mining where data is processed to give information and visualization of interesting patterns found in the user web log data, are performed. In this step of Web Usage mining the insignificant rules are removed, data is represented and then OLAP tools are applied. Pattern discovery and analysis are major processing phases of web usage mining. [6]

3. WEB USAGE MINING TOOLS

3.1 Webminer

WEBMINER is a general architecture for web usage mining process[2]. In this architecture, the process of web usage mining is split up into 2 main section. The first section included the domain dependent process for transform web data into suitable transaction form. This step of processing includes preprocessing, transaction identification and data integration components. The second section included largely domain independent application of generic data mining and pattern matching techniques.

3.2 WEBKIV

WEBKIV included mainly four basic components based on its architecture

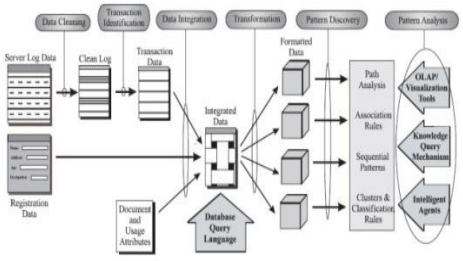

Figure 1:Architecture of WEBMINER [2]

and the process of visualization in WEBKIV is described further[1].

WEBKIV is a web mining tool developed to visualization of results of web server log in WEBKIV, which combined the tasks from other visualization tools so that provide a single technique of visualization of data structure, and it deploys the results on the structure. WEBKIV stands for Web Knowledge and Information Visualization designed to experiment with visualization of structure content and navigation.

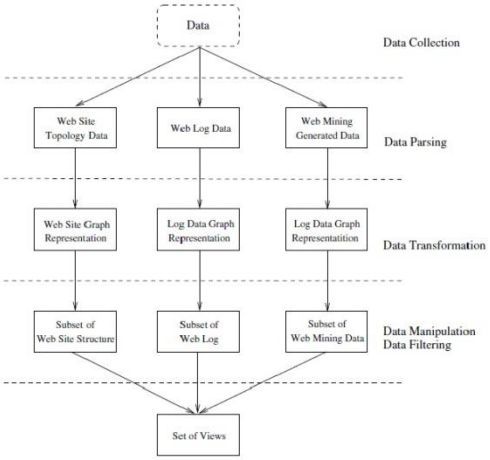

Figure 2: **Architecture of WEBKIV**[1]

The main functions of WEBKIV are as follows:

- Web structure visualization: Provide tool for visualize web structure (small and large), with controls that support both type of structure detailed and abstract.
- Web navigation visualization: Here tool provides static visualization and dynamic display.[1].
- Web mining results comparison: WEBKIV tool uses overlaying web navigation patterns, and comparing those constructed from the application of machine learning to navigation improvement.

3.3 WebSift

WebSift[18] is a system based on the WEBMINER prototype[2] used the information and content from the particular website. This is the framework based upon web usage mining access the clustering, content and duration of the process mining. WebSift is used to identify the potential new interesting results from the log data.

3.4 RAPIDMINER

RAPIDMINER is a web usage mining tool used basically for business analytics, data mining basically based over machine learning, acts as client-server model released in 2006. It uses XML to describe

operator tree and easily read excel format and different types of formats

4. COMPARISON OF WEB USAGE MINING TOOLS

Here we present the comparison of these tools of web usage mining and identify the performance of these tools. Here compare these tools, namely, WEBMINER, WEBKIV,WEBSIFT and RAPIDMINER on the basis of performance parameters.

Techniques and Performance Parameter	WEBMINER [2]	WEBKIV [1]	WEBSIFT [18]	RAPIDMINER [4]
CLUSTERING[4,2,18]	□		□	□
ACCURACY[4]	□		□	□
NUMBER OF PAGES/ITERATION[4]	□		□	□
TIME TAKEN[4]	□		□	□
ERROR RATE[4]	□		□	□
SEQUENTIAL ACCESS PATTERN[4,1]	□	□		□
CLASSIFICATION[4]	□			□
ACCURACY[4]	□			□
MEAN ABSOLUTE ERROR				□
ROOT RELATIVE ERROR				□
PRECISION[4]	□		□	□
RECALL				□
FMEASURE				□
MCC				
TP RATE[4]				□
FP RATE[4]				□
PRC AREA				□
ROC AREA				□
ASSOCIATION RULE[4,1,18]	□	□	□	□
ACCURACY	□		□	□
TIME TAKEN	□		□	□
NUMBER OF FREQUENT ITEMS	□		□	□
REGRESSION				□
AGGREGATION[1]		□		□
VISUALIZATION [4,2,1]	□	□		□
1-D,2-D(X-AXIS, Y-AXIS)	□	□		□
3-D(Z-AXIS)				
TREE MAP		□		□
COLOR CODE		□		□
PLOT SIZE	□	□		□
POINT SIZE	□	□		□
JITTER				
ZOOM		□		□

FEATURE OPTIMISED				□
TIME TAKEN[4,1]		□		□
CONTENT[4]				□
FREQUENCY[18]			□	□
FAILURE REQUEST	□			□
DURATION[4,1]	□	□	□	□

TABLE 1

In Table 1 we present the comparison of these three tools on the basis of techniques they use. The major techniques used in all these tools are Preprocessing, Sequential access patterns, Clustering, Classification and Visualization of the web log data that we fetch from servers accessed by users.

5. LIMITATIONS

WEBMINER is a prototype tool[18] for web usage mining not available for user, secondly WEBMINER is concern for association rule there is not much source available for classification of data as only C4.5 algorithm[2] is only available in it and second it process a very low as gap in research or invented very earlier. As in classification it is only calculated precision while WEBKIV is access only for association rule based on WEBMINER prototype have no classification for data. WebSift is a web usage mining tool based on WEBMINER have only option for clustering but no further improvement is not seen in these described tool.

RAPIDMINER[19] is a new tool compared with these previous tools and have more option than WEBMINER,WEBKIV and WEBSIFT but RAPIDMINER is not suited for classification of data as it process the rules of WEKA. RAPIDMINER is a tool best suited for people who worked on database files second in RAPIDMINER to visualize the data not much facility available not supported the 3-D graph.

6. WEKA

WEKA is a powerful tool used for mining data from online web data or offline data. In WEKA we can use machine learning classifiers on data mining tasks. WEKA contains tools for preprocessing, regression, association, clustering, classification and visualizing web data. These techniques are applied on web data using code written in Java[3].

7. COMPARISON OF WEBMINER, WEBKIV WITH WEKA

Here in table 2 we presented the analysis report of WEBMINER, WEBKIV and weka tool comparison on basis of techniques they follow. After analysis out of these tools we stated that weka give more better result on basis of performance on web server log data.

Technique	**WEBMINER [4]**	**WEBKIV [1]**	**WEKA [3]**
Association rule and sequential pattern generation	Available for data mining algorithms(Apriori Algorithm only)	association rule is implemented with limited resources use only breath first search algorithm for parsing	Provide more facility to analyze the data (Apriori,FP Growth)
Dataset	Cannot handle larger dataset	used dataset based on session. No variety of access different dataset	Can handle large set of data
CLASSIFICATION	Limited number of classification technique(C4.5 ALGORITHM)	No availability of classification algorithms	large set of classification algorithms technique available

Pattern Visualization	Rigorous analysis for transaction identification approach.	single type of visualization is available. Not much variety for different pattern visualization	Deep information can be analyzed
Regression Analysis	No option for Regression Analysis	not available	Regression analysis available
Filteration	Limited filtration facilities	feature not available	Large set of filtration of dataset is possible
Server log files	Limited access to log files in system	used only static pages no more option for auxiliry pages ins erver log files	Access to log files of every kind in today world while WEBMINER have access restricted number of log files

TABLE 2

8. CONCLUSION AND FUTURE WORK

In this paper we presented major web usage mining tools. The first, namely, WEBMINER, is a framework depicting the working of web usage mining in three phases, that is, pre-processing the data, pattern discovery to discover new patterns in web log data, and, lastly, pattern analysis, which is used to identify new interesting patterns of the data. Thereafter, WEBKIV is used to visualize the data in the form of plots. The third important tool weka, which works on the KDD process as compared with both the tools stated above.

WEBMINER is a web usage mining tool invented by [4] brief a outline only how web usage mining process but WEKA work it commercially work same as WEBMINER. WEBMINER system not grow up with time while WEKA changes in their system time to time and make it suitable for now a days large set of database of every kind. While WEBKIV system is limited only to association rule changes and visualize the data but in WEKA provides a variety of different approach of today latest technology as we compared with the both previous tools. Both tools WEBMINER and WEKA work for analysis of data and generation of new rules over the web but WEKA provide more facilities and optimized data when we compared it with WEBMINER.

From the Table above and the techniques available in these tree tools at last we stated that WEKA gives better performance results when compared overall with WEBMINER and WEBKIV. In WEKA we look that in every technique like sequential pattern, association rule, clustering and classification WEKA have more features with optimized algorithms which not present in the WEBMINER and WEBKIV. Thus we concluded that the performance factor of WEKA is better than WEBMINER and WEBKIV and so that the result of WEKA is more optimized than other two tools i.e. WEBMINER and WEBKIV

REFERENCES

[1] NiuY.,ZhengT.,ChenZ.,Goebel R.,(2003)"WEBKIV: Visualizing Structure and Navigation for Web Mining Applications" at Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003), Year: 2003

[2] SrivastavJ.,Cooley R., Mobasher B.,(1997)" Web Mining Information and Pattern Discovery on world wide web" at Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence,1997,**ISBN:** 0-8186-8203-5

[3] Kulkarni G.,E.,Kulkarni B. Raj,(2016)"WEKA Powerful Tool in Data Mining", International Journal of Computer Applications (0975 – 8887) National Seminar on Recent Trends in Data Mining (RTDM 2016)

[4] Renáta I, IstvánV., (2006) ,"Frequent Pattern Mining in Web Log Data",at Acta Polytechnica Hungarica Vol. 3, No. 1, 2006

[5] Chavda S., Jain S. Valera M.,(2017)" Recent Trends and Novel Approaches in Web Usage Mining", at International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 | April -2017 e-ISSN: 2395 -0056

[6] V.Anitha,Dr.P.Isakki,(2016) "A Survey on Predicting User Behavior Based on Web Server

Log Files in a Web Usage Mining" at 978-1-4673-8437-7/16/$31.00 ©2016 IEEE

[7] Sakthipriya C., Srinaganya G., Dr.Sathiaseelan J.G.R.,(2015)" An Analysis of Recent Trends and Challenges in Web Usage Mining Applications", at IJCSMC, Vol. 4, Issue. 4, April 2015, pg.41 – 48

[8] Shoaib M.,Maurya A.,(2014). " Comparative Study of Different Web Mining Algorithms to Discover Knowledge on the Web", at Proceedings of the Second International Conference on "Emerging Research in Computing , Information, Communication and Applications" ERCICA 2014, ISBN 9789351072638

[9] BoullosaJ., XexéoG." An Architecture for Web Usage Mining", at https://www.researchgate.net/publication/228409264_An_Architecture_for_Web_Usage_Mining

[10] Ms.Aparna M. Katekarat,(2017)"Improving the Effectiveness of Short Text Understanding by Using Web Information Mining" at Proceedings of the IEE, International Conference on Computing Methodologies and Communication(ICCMC), 2017

[11] BrijendraSingh, Hemant Kumar Singh,(2010) "WEB DATA MINING RESEARCH: A SURVEY" at 978-1-4244-5967-4/10/$26.00 ©2010 IEEE

[12] V.Chitraa, Dr. Antony SelvdossDavamani,(2010) "A Survey on Preprocessing Methods for Web Usage Data" at (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 3, 2010

[13] Aditya S P, Hemanth M, Lakshmikanth C K, Suneetha K R,(2017) "Effective Algorithm for Frequent Pattern Mining"at International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017),2017

[14] M.Sathya,Dr.P.Isakki@Deviat,(2017)"Apriori Algorithm on Web Logs for MiningFrequent Link" atIEEE International Conference On Intelligent Techniques In Control , Optimization and Signal Processing,2017

[15] Dr.K.Shyamala, S.Kalaivani,(2017)"An Effective Web Page Reorganization throughHeap Tree and Farthest First ClusteringApproach"atIEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017)

[16] SunHao, ShenZhaoxiang, ZhangBingbing,(2017)"A User Clustering Algorithm on Web Usage Mining" atFirst International Conference on Electronics Instrumentation & Information Systems (EIIS),2017

[17] Cooley, R., Tan, P-N, & Srivastava, J (1999). Web SIFT: The Web Site Information Filter System

[18] Rangra K.,Dr Bansal K.L.,(2014). "Comparative study of Data Mining Tools" at International Journal of Advanced Research in Computer Science and Software Engineering