Recognition of Interested Web Users Behavior

Sadhna K. Mishra
Professor, Dept.of Computer
Science
LNCT Bhopal

Vineet Richaria
HOD, Dept of Computer
Science
LNCT Bhopal

Vivek Sharma
M.Tech.(Software Engineering)
LNCT Bhopal

ABSTRACT

There are some words like web browser, web server, web applications and websites, have become buzzwords. The search engine like google have become so much ubiquitous that everyone want to search the information on google .The web mining is a technology that is provide easy technique to use the web applications. The web application developer has the prerequisite of analysis of user's behavior for the development of any web application. The analysis of user behavior show who are the interested user and not interested user. There have been used a classification algorithm based on decision table. The Deep Log Analyzer tool has been used for the analysis of user's behavior. This analysis will help to developer to develop the web applications according to requirements of web users. This analysis is similar to case study of web user's behavior.

General Terms

Classification, Log Analyzer, Web User Behavior

Keywords

Web Usage Mining, Deep Log Analyzer, Decision Table.

1. INTRODUCTION

The web mining has become a growing and emerging field for research in computer science. This is due to the internet users are growing in leaps and bounds or in exponential manner. The internet are using for searching the information of every discipline. The internet has the prerequisite of search engine, web browser, web server and database to store and retrieve the information. The tracking tools of web user behavior use the ip address, session, browser, operating system, no. of pages, depth of pages, and path of pages and objects like sports, fashion, education and business etc..The information is collected from the tracking tools stored in the log files. The log files may be in the CVV file format or in the ARFF file format. These log files are stored on the server. The log analyzer tools like AWSTATS, DEEP LOG ANALYZER, and WEKA TOOL etc. are used for the analysis of user's behavior. We have used here The Deep Log Analyzer Free version tools and sample data available with this tool.

2. RELATED WORK

There have been done a lot of work in the research analysis on user's behaviors. There has been used different type of classification methods and characteristics for segregating to interested user and not interested users.

We have used here Classification based on decision table; in this paper C4.5 algorithm are used and attributes are selected from a dataset [7]. There have been already done a work based on decision tree ,in this paper also the C4.5 have been used for the attribute selection [8].The B.Santosh kumar and Rukumani have used APRIORI and FP Growth algorithms to web usage mining, in the Apriori algorithm ,there have been searched the different association rules between frequent items. The FP growth algorithm based on frequent item pattern generate by FP tree structure [3].The Jaideep Srivastava and Robert Cooley have Been searched the different pattern of web users behaviors and explained the tool WEBSIFT [5].In this survey, the different techniques and research issues of web usage mining has been explained, there are focused particularly on web content and user's behavior [1]. R.suguna has done work on overview of web usage mining and research work so far in this area [2]. The user's profiles have been prepared by the tracking and analysis of user's behaviors. It has performed for the customer relationship management .This profiles include both academic and industrial efforts, as well as Commercial offerings [6].

3. CLASSIFICATION

There have been used so many classification methods to identify the users pattern. The pattern may be based on Association rules, frequent item sets, page reference and decision tree. The decision tree work on attributes selection from the files such as arff (attribute related file format) files. The attributes may be session, ip address, and page reference, no. of visitors and depth of visitors.

In this paper, we have used the classification of web user based on decision table. The decision table provides the multipath of selection. There have been used global root nod of decision table and local root nod of decision tree. At the global root node we can select the popular web sites based on number of visitors that site and at the local root node we can search the pattern of web user's behaviors on particular website. The decision table work on the base of constraint rules such as if visitors no is more than hundred for a particular website then that website will be analyzed as a popular website. If visitor's number is less than hundred then that web site is less popular. Here P1…………P6 are pages.
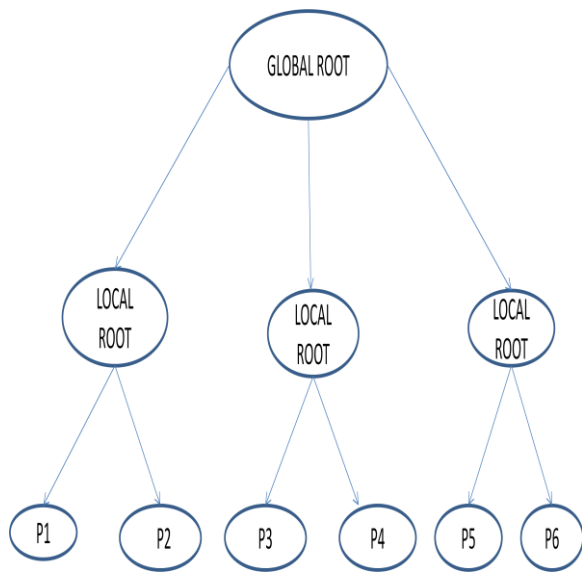
Fig.1 Decision Tree

4. ANALYSIS

There have been done a lot of work in the field of user's behavior analysis so far. But we have used here decision table for the first time to analyze to users behavior. The Decision table is a structure tool for the designing phase of any software development phases. The designing have been a greater part of cost of the software development because it represents the user's requirement visually.

There have been used many tools for the recognition of users behaviors such as WEKA tool and AWSTATS tool. There have been developed mostly tools in PHP language but WEKA tool has been developed in JAVA. We have used Deep Log Analyzer Free Tool for the analysis of web users behaviors. The Deep Log analyzer Tool has the sample data in MS Access file format. The deep log analyzer also has the tracking facility of user's detail.

The Deep log analyzer tool has tracked the data of following website http://www.interactivegt.com.

The data have been uploaded to the web server in the MS access file format. The report date interval has been given 14 days. There have been tracked many things about the user, Websites, depth of visits, number of visits, page views, search engine summary and referral summary of referred page. We have select referring sites; visitors stay length, number of visits per visitors and depth of visitors and concluded in the decision Table.

5. RESULTS

The following figure shows popular sites, number of visits by visitors, visitors stays length and depth of visit. These are the different attributes to analysis and the conclusion by these attributes has shown in the last table that who is the interested user and not interested user.

The referring sites given below in the figure 2 provide the information that which sites mostly used. We can see here that w3school is most popular site rather than other sites.
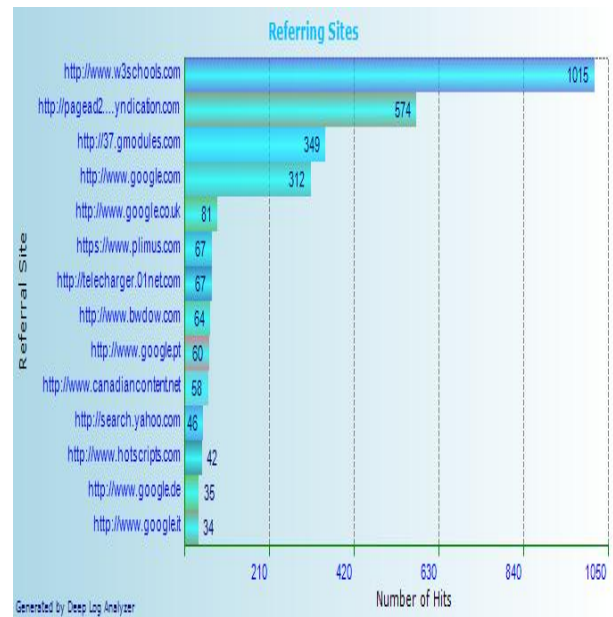

Fig.2 Referring Sites

The top pages available on different sites are another attributes that provides the information about the pages accessed more rather than other pages. We can conclude that which page has view more rather than other pages.
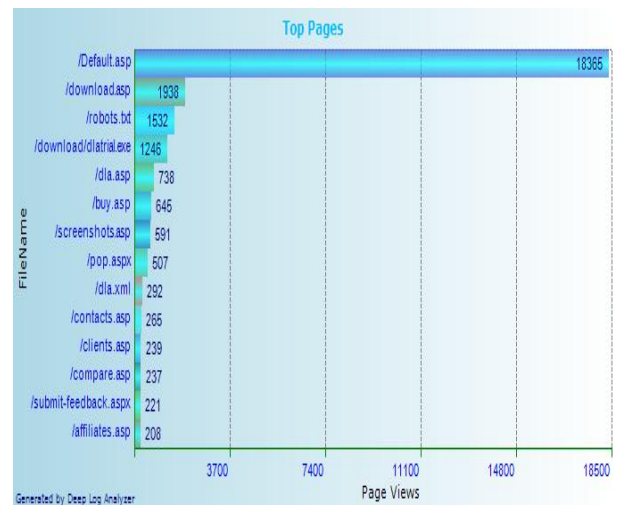

Fig.3 Top Pages

The figure.4 explains the detail about the visitors who have visits the sites more times rather than other visitors. It will provide the information about user who has more interested and not interested. We can analyze from the figure that some visitors visit only once those are not interested and some visitors visits more than 51-100 times those are most interested users.
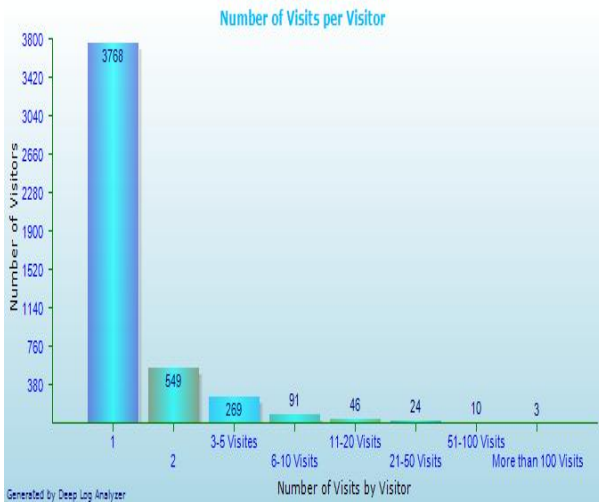
Fig.4 Number of Visits per Visitor

The following graph provides the information about the users who access the more than one page in a web site. The users who visits only one page is not interested users rather than who visits more than 30 pages.
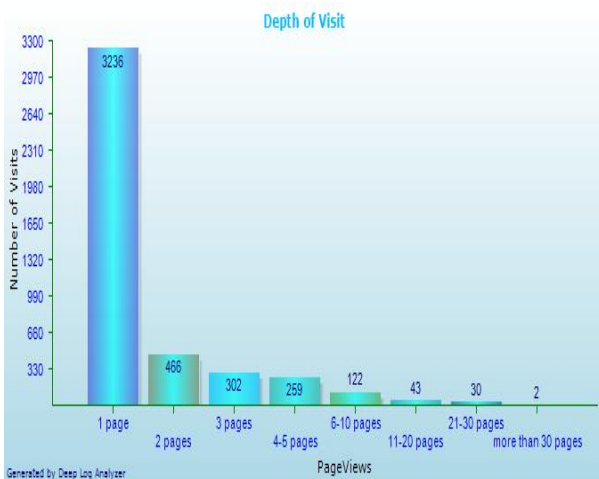

Fig.5 Depth of Visit

The following graph provides the visitors stay length**.** The visitors stay length show that the more length is indication for interested users.

Fig.6 Visitor stay Length

The decision table has two section .one section is condition stub and others is action stub. There have been given different condition in the form of attributes achieve from the above graph analysis. The actions are taken according the rules R1, R2, R3, and R4 and conditions. If condition match with R1 and R2 then action will taken that the user are interested and else if match with R3 and R4 then the users will not be interested users

.Here

IU = Interested Users

NIU = Not Interested Users

Conclusions:

There have been used decision table for the classification of interested users and not interested users. The decision table provides the multipath for the selection of popular sites and attributes. The decision of interested users can help to develop web sites according to user's requirements .The internet demand is growing in leaps and bounds and the internet is a repository of information. The information is stored in the web sites. The designing of web sites according users requirement is very challenging. The information about interested user would great help to design the web sites. There is billions of users search the information on internet about various discipline. The log analysis tool provides the analysis about the users that which user is interested in which type of information. These tools help to web site developer to design semantic of web sites. The web sites semantic provide a structure of web sites in which users can access information very easy way. The classification techniques of information arrange the information in various classes depend on attributes. There is different method to classify the information. The classification method based on decision table is very easy to classify. There is different analysis method based on different classification technique. The analysis based on decision table provides the information about users those are interested in particular web sites.

Future Work:

There are many tools available for log file analyze. There tools are developed in PHP, but there are not available tools in java except WEKA tools. The WEKA tools use ARFF files for web data. The future work will be focused on java tool that will provide more security and platform independent environment. The new tool will provide the facility of accessing every type of files of web data. The java tools provide the more security because it uses the JVM. The java provides the various classes for classification. Java also provides the tree and table classes designed by the sun micro system. The JSP (java server pages) is uses for the development of dynamic pages. There are requirement of very less coding. The Struts provide the user friendly frame work to develop any web sites. The Hibernate provides the facility of connection with any database of applications.

Fig.7 Decision Table

	R1	R2	R3	R4
Popular Site >630	Y	N	N	N
Popular Site <630	N	Y	Y	Y
Depth of Visit>10	Y	Y	N	N
Depth of Visit<10	N	N	Y	Y
Number of Visits per Visitiors>20	Y	Y	Y	N
Number of Visits per Visitiors<20	N	N	N	Y
Visitiors Stay Length>20	Y	Y	Y	N
Visitiors Stay Length<20	N	N	N	Y
Action	IU	IU	NIU	NIU

The rules are the following

R1   Popular path through site> average

Depth of visit>average

Visitors stay length>average

No. of visits per vistors>average

R2   Popular path through site< average

Depth of visit>average

Visitors stay length>average

No. of visits per vistors>average

R3   Popular path through site> average

Depth of visit<average

Visitors stay length<average

No. of visits per vistors<average

R4   Popular path through site< average

Depth of visit<average

Visitors stay length<average

No. of visits per vistors<average

6. References

[1] Rakesh Kumar Malviya, Mahesh Chandra Malviya, Vinay Kumar Soni, Ritesh Joshi,and Preetesh Purohit," Survey of Web usage Mining," IJCST vol. 2,Issue 3,September 2011.

[2] R.Sugunal, D. Sharmila, An Overview of Web Usage Mining.IJCA(0975-8887) vol. 13-No.13, Febuary 2012.

[3] B.Santosh kumar, and K. V. Rukmani, "Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms," in IJANA, vol. 01, Issue:06 Pages: 400-404(2012).

[4] S.Taherizadeh, N.Mooghadam, "Integrating Web Content Mining into web Usage Mining for Finding Patterns and Predicting Users' Behaviors" IJISM,Vol. 7,No. 1(2009)

[5] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan,"Web Usage Mining: Discovery and Applicatons of Usage Patterns from Web Data" SIGKDD Explorations,ACM ,Jan 2000.

[6] Amsaveni. K, Vydehi.S,"A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites" IJCTT-volume3 Issue4-2012.

[7] Sadhna Mishra, Vivek Sharma, "Classification of Web Users into Interetsted Users and Not Interested Users By Using Decision Table" IJARCS volume 3 No. 4. 2012.

[8] K. R. Suneetha, R. Krishnamurthy. Bharathidasan , "Classification of *Web* Log Data to Identify *Interested Users Using Decision Trees" IJCSI* International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

[9] Robert Cooley, B. Mobster, and Srivastava: Data preparation for Mining World Wide Web Browsing Patterns, Knowledge and Information System., Volume 1, No 1, pp. 5-32 (1997).