ISSN No. 0976-5697

Volume 3, No. 4, July- August 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Classification of Web Users into Interested Users and Not Interested Users by Using Decision Table

Prof.Sadhna Mishra
Dept.of computer science
LNCT ,Bhopal(M.P.)
sadhnaguddy@rediffmail.com

Vivek Sharma*
M.Tech.(software engineer)
LNCT,Bhopal(M.P.)
vivek_gwaliormp@yahoo.com

Abstract: It is an era of information. There are bulks of information available on web sites. The mining of such a huge amount of data is a critical job. Web mining provides easy technique to search web data. The data is available on web site may be or not be useful for web users. Hence data is classified into interested and not interested data according to useful and not, for users. Here we are using decision table for classification of web data. The algorithm C4.5 is used in this paper. There are three phases in web usage mining preprocessing, pattern discovery and pattern analysis. In the first phase we will process data for the segregation of useful and not useful data. The second phase will carry out for the discovery of pattern in which user search the data and the third phase could be used for pattern analysis for designing the web sites.

Keywords: web usage mining, decision table, web data.

I. INTRODUCTION

The web mining is an essential technique for those people who are searching information. It has become a buzzword for all the technical and not technical people. The web usage mining is a part of web mining that is again a part of data mining. Because the selection of appropriate data is very tedious job from giant amount of data .Hence first of all we will perform preprocessing the information by using C4.5 algorithm. Then we will categorize information by using decision table. The decision table provides the multipath in comparison of the decision tree. The categorization is based on the selection of attributes. First of all we will find out the entropy of the attributes. The Entropy is difference between the attributes. Then we will find out the information gain The Information gain is maximum when entropy is minimum. We are using here C4.5 algorithm for the splitting of attributes. Here, we will use msnbc.com data sets for analysis of the user's pattern

II. RELATED WORK

The web usage mining is new emerging area for research although there have been done a lot of works but still there are lots of possibilities to research in this area.

The paper [1] present the classification of interested user based on algorithm C4.5 and [2] Cooley and jaideep shrivastav has described about the web usage mining preprocessing, pattern and analysis of web data. In the link [3] has given web data on the web site msnbc.com and the paper [4] describe the interested users by Naive Bayesian Classification. Angha shastri, Dipti Patil and Wadhni has analyzed the user behavior based on constraint in their paper [5].

The jean Pierre and Benjamin has done classification based on the paper [6]. S.Taherizadeh and N. moghadem has published their paper[7] on finding pattern and prediction of user behavior.Rajni pamnani and Pramila chawah has performed the analysis of web usage mining in their paper[8]. Krithuka and Mrs.Dipa dixit has done work on pattern discovery in web usage mining[9].

There are many research work focus on the decision tree for classification. The work has completed by using C4.5 and ID3 algorithm. We are here focused on the decision table because it provide multipath concurrently.

III. DECISION TABLE

The decision table is a structured analysis tool .These tools are used for the designing purpose. The decision table has been used for the decision purpose. We have used C4.5 algorithm for designing the decision table although C4.5 algorithm is used for the designing of decision tree. We are using table because it provide multipath concurrently. The decision table structure has given below:

Table 1: decision table

	Rules
Condition stub	
Action Stub	

IV. CONSTRUCTION OF DECISION TABLE

The construction of decision table will be initialized by the attributes. The attributes are the following

Table 2: Atrributes

Attributes	Description
Session Time(S1)	Total time to searching the web sites
Session Time(S2)	Total time to searching a website

No. of Pages(P)	Number of searching pages
Method	Access methods
Depth	Access according to depth

The conditions for the decision table based on these attributes

Table 3: conditions

Conditions	Descriptions
1	S1>30minutes
2	S2>10minutes
3	S2<10
4	P>5
5	P<5
6	Get
7	Post
8	Yes
9	No

Now the table would be designed on the basis of these conditions .Here R1……..R12 are rules.

Table 4: Decision Table

	R1	R2	R3	R4	R5	R6
S1>30 S2>10 S2<10 P>5 P<5 GET POST YES NO	Y	Y	Y	Y	Y	Y
	Y	Y	Y	Y	Y	Y
	N	N	N	N	N	N
	Y	Y	Y	N	N	N
	N	N	N	Y	Y	Y
	Y	Y	N	Y	Y	N
	N	N	Y	N	N	Y
	Y	N	N	Y	N	N
	N	Y	N	N	Y	N
Action	IU	NIU	IU	IU	NIU	IU

	R7	R8	R9	R10	R11	R12
S1>30 S2>10 S2<10 P>5 P<5 GET POST YES NO	Y	Y	Y	Y	Y	Y
	N	N	N	N	N	N
	Y	Y	Y	Y	Y	Y
	Y	Y	Y	N	N	N
	N	N	N	Y	Y	Y
	Y	Y	N	Y	Y	N
	N	N	Y	N	N	Y
	Y	N	N	Y	N	N
	N	Y	N	N	Y	N
Action	IU	NIU	IU	IU	NIU	IU

Here
IU=Interested User
NIU=Not Interested User

There are two types of users those registered or not registered. The registered users use the post method and they are obviously interested users. The not registered users use get method. They may be interested or not interested decide by the depth searching by them. The decision table could be used for the multipath .By the multipath facility we can analyzed concurrently different type of web sites user. Here we can differentiate user by S1>30

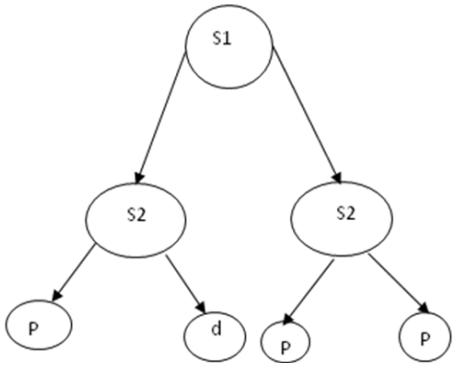

Figure 1: Multipath tree

V. CONCLUSION

There have been focused on the interested users in this paper. This work could be help to design the web sites only focused on interested users. It will save not only time and cost of developer but also reduce the searching time of user. The decision table provide the multipath hence it is possible to the different websites analysis searching behavior of user.

VI. REFERENCES

[1]. K. R. Suneetha, R. Krishnamurthy. Bharathidasan ... Classification of web *log data* to *identify*. Interested users using decision trees ijcsi *i*nternational Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012

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

[3]. http://kdd.ics.uci.edu/databases/msnbc/msnbc.html

[4]. A. K. Santra, S. Jayasudha Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012 ISSN (Online): 1694-0814 www.IJCSI.org

[5]. Anagha Shastri,Assistant Professor, BIT, Mumbai india, Dipti Patil,Assistant Professor,MITCOE,Pune, IndiaV.M. Wadhai,Professor and Dean of Research,MITSOT, MAE, Pune, , India Constraint-based Web Log Mining for Analyzing Customers' Behavior International Journal of Computer Applications (0975 – 8887) Volume 11– No.10, December 2010

[6]. Jean-Pierre Norguet1, Benjamin Tshibasu-Kabeya2, Gianluca Bontempi2, Esteban Zim¶anyi1 A Page-Classification Approach to Web Usage Semantic Analysis Engineering Letters 14:1, EL_14_1_21 (Advance online publication: 12 February 2007)

[7]. S. Taherizadeh N. Moghadam, Integrating Web Content Mining into Web Usage Mining for Finding Patterns and Predicting Users' Behaviors, Integrating Web Content Mining into Web Usage Mining for Finding Patterns and Predicting Users' Behaviors International Journal of

Information Science and Management, Vol. 7, No. 1 January / June, 2009

[8]. Rajni n Pamnani, Pramila Chawan, "Web Usage Mining: A research area in Web Mining", Punjab, ISBN 978-81-910304-0-2 International Conference on Recent Trends in Computer Engineering, ISCET 2010, RIMT,

[9]. Mrs. KIRUTHIKA M, Mrs. DIPA DIXIT, MINING ACCESS PATTERNS USING CLASSIFICATION, Kiruthika M et. al. / International Journal of Engineering Science and Technology Vol. 2(7), 2010, 2895-2902