ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 1, January 2018

Web Page Navigation Analysis from Log File in Web Usage Mining

[1] Divya Patel, [2] Hetaxi Kamli
[1] Assistant Professor, [2] Software Engineer

Abstract: **- Web mining is the process of finding web pattern from web data. Web link mining finding useful things from log file. Support vector machine is used to classify all the data of web log file that analyze data and identify patterns. After applying SVM, Data into one pattern can be discovered. This pattern is useful for business analyst to take a useful decision. Finally we get classified data from web user navigation data.**

Keywords: - Appriori, SVM.

I. INTRODUCTION

Data mining is the method of finding useable patterns or knowledge from data sources such as databases, web log etc. Web mining gives information which is produces form web data. It is mainly divided into Web Content, Web Structure and Web Usage information. Web mining mainly searches for the hyperlinks and the movement pattern of user. Web use mining is fully focused on process that utilized to predict the client behavior while client interact with the web site. Web usage mining is the method of focusing handy data from server logs for example client's history.

II. WEB USAGE MINING

Web usage mining uses data mining techniques to find out valuable data from navigation pattern of Web users. The primary data is collected by servers and stored in Web logs. The first task in Web usage mining is the pre processing of data. In preprocessing stage, first related data is filtered out from the web logs.[1]

There are mainly three processing step in Web Usage Mining.[2]

A. Preprocessing:

It contains mainly three phases:

(1) Cleaning: in this phase useless and unwanted entries removed.

(2) Session identification: in this phase each unique session is created which has all request of one user.

(3) Data conversion: in this phase data is converted into the specific format of particular software tool.

B. Pattern discovery

It means finding pattern of user interaction with web site by applying the mining algorithm which has user's predefined constraints.

C. Pattern analysis

In this task human comes into the action. Human has to understand the domain of system and on the basis of that domain analysis can be done to filter out useful and understandable pattern and drawing conclusions.

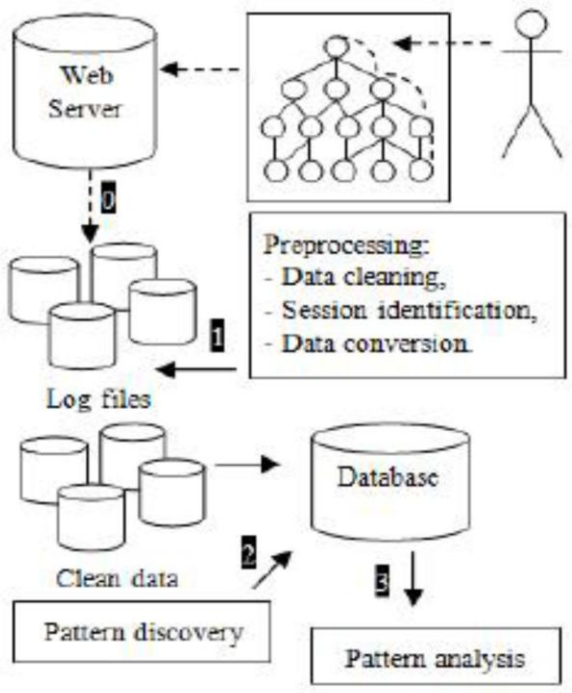

Figure 1: Logical view of Web usage mining [1]

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 1, January 2018

III. PROPOSED WORK

SVM is based on machine learning. It is known for the superior classification result.[3]

A. Flow of Propose Algorithm

Steps of Proposed Algorithm

(1) Input Log file: In this step, select one server file form which select log file on behalf of different parameter.

(2) Feature Selection: In this step, select data as per user or other parameter like on the base of IP address, Machine base etc.

(3) Support Vector Classification: After selecting data from the user log file, classify data into different classes. Here two classes are separated by hyper plane.

(4) Classification Result: After classification we get result which is divided into two different classes.

(5) Get Frequent Items: By using support and confidence, find the frequently appearing items in dataset.

(6) Generate Candidate Item sets: After finding frequent items, generate candidate item sets.

(7) Get Frequent Item sets: After candidate items, generate frequent item sets which fulfill the condition of minimum support and minimum confidence.

(8) Generate Strong Rules: By generating frequent item sets, strong rules can be created for frequent pattern mining.

B. Flow of Proposed Algorithm

The flow of proposed algorithm is shown below figure.

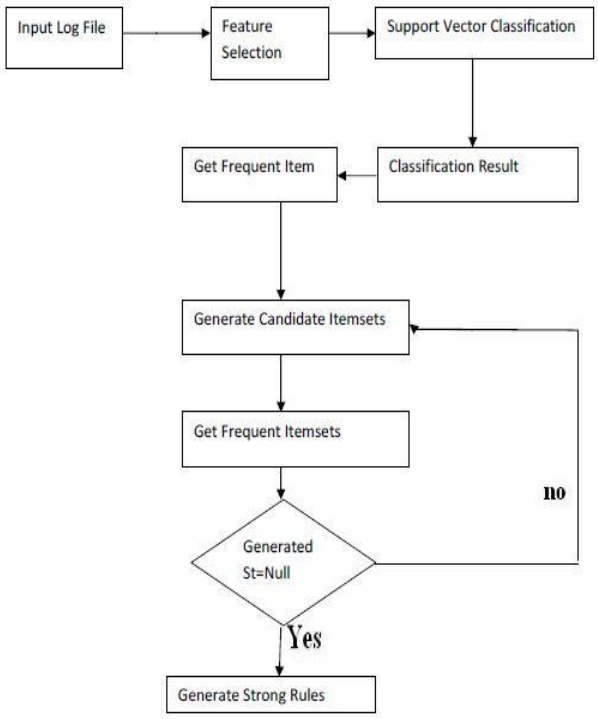

Figure 2: Flow of Proposed Algorithm

IV. RESULT

As shown in above image proposed algorithm work well. It eliminates itemset and still gives accurate output. Following graph shows how many number of candidate set is reduce. On y axis it show number of candidate set and on X axis Minimum support is given.

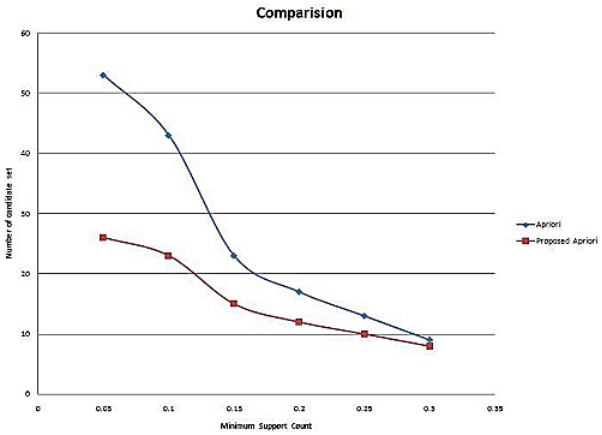

Figure 3: Graph of comparison

V. CONCLUSION

The primary goal is to extract information from server web log file i.e. in semi structured format, and transforms this information into meaningful information for instance there is framework suggested by us and implementation also provided for utilization of extracted information in prediction of next access. Apriori only filter small amount of data. So execution time increase but using SVM we can filter large amount of data without affecting execution quality. In SVM classification technique we can classify data into different classes which give us a batter view of data. Using SVM classification technique we can classify data using different criteria on same data. At the end execution time is less and lot of processing power can be saved.

REFERENCES

[1] Paweł Weichbroth, Mieczysław Owoc," Web User Navigation Patterns Discovery from WWW Server Log Files", 978-83-60810-48-4, 2012 IEEE.

[2] Mr. Akshay A. Adsod, Prof. Nitin R. Chopde, "A Review on: Web Mining Techniques", ISSN: 2231-5381 IJETT – Volume 10 Number 3 - Apr 2014.

[3] S. Chitra, B. Kalpana , " Transductive Support Vector Machine Web Log Classifier for Identifying Potential Users ", ISSN No: 2309-4893 Vol-2, Issue-7, July 2014.

[4] Aarti Sharma, Rahul Sharma, Vivek Kr. Sharma, Vishal Shrivatava, "Application of Data Mining – A Survey Paper" International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2023- 2025

[5] Sawan Bhawsar, Kshitij Pathak, Sourabh Mariya, Sunil Parihar, "Extraction of Business Rules from Web logs to Improve Web Usage Mining", ISSN 2250- 2459, Volume 2, Issue 8, August 2012.

[6] Jianli Duan, Shuxia Liu," Research on web log mining analysis", 2012 IEEE.

[7] Theint Theint Aye"Web Log Cleaning for Mining of Web Usage Patterns", 978-1-61284-840-2/11, 2011 IEEE.

[8] J Vellingiri, Dr.S.Chenthur Pandian, "A Survey on Web Usage Mining" Global Journal of Computer Science and Technology Volume 11 Issue 4 Version 1.0 March 2011

[9] Anagha Shastri, Dipti Patil, V.M.Wadhai, "Constraint-based Web Log Mining for Analyzing Customer's Behavior", International Journal of Computer Applications (0975 – 8887) Volume 11– No.10, December 2010.

[10] Renáta Iváncsy, István Vajk, "Frequent Pattern Mining in Web Log Data", Vol. 3, No. 1, 2006

BOOK REFERENCES

[11] Jiawei Han and Micheline Kamber "Data Mining Concepts and Techniques", Morgan Kaufman publishers, San Francisco, Elsevier, 201, pp. 285-351.

[12] Bing Liu "Web DataMining: Exploring Hyperlinks, Contents, and Usage Data", Springer-Verlag Berlin Heidelberg 2007