International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 1 | January -2017 www.irjet.net p-ISSN: 2395-0072

Classification of User & Pattern discovery in WUM: A Survey

Shahu P.S.[1], Jamthe D.V.[2] , Nikose A. A.[3]

[1]M.Tech Student [1]Dept. of CSE, PBCOE, Maharashtra, India

[2]Asst. Prof Dept. of CSE, PBCOE, Maharashtra, India

[3]Asst. Prof Dept. of CSE , PBCOE, Maharashtra ,India

Abstract - *Web usage mining effectively and efficiently serves the needs of the user visiting the websites. Web usage mining is based upon the discovery and analysis of web usage patterns from web logs. The purpose of this survey to identify the user's classifications based on discovery pattern from web logs. The work is done in three steps. Preprocessing is done in first step to remove useless data from web log file that reduced its size. In next step, for discovering usage pattern cleaned log file is uses. In last step discovered pattern leads to the classification of user. Classification of users is identified on the basis of countries, direct entry to the site or referred by the other site, time of access, user and IP address based. All the identifies classification of user can be used by the administrator for the effective administration and website personalization. That result into fulfill the specific needs of specific community users and hence the profit can be increased.*

***Key Words*: Web mining, Web usages mining, Pattern discovery, SVM.**

1. INTRODUCTION

Web mining is the process of examining data sets collected from various sources methodically and in detail, in order interprets it to get useful information. These data sets may consist of web log data. Researchers have classified web mining into 3 types, namely, web structure, content and usage mining. This classification is based on the type of data to be mined [5].

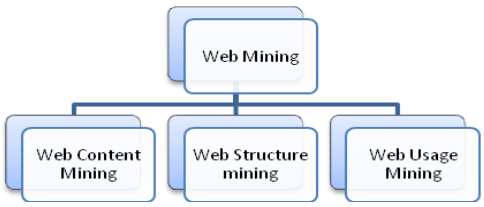

Fig 1: Classification of Web Mining

Web usage mining is also termed as web log mining. It is based on discovery and analysis of web usage pattern from web logs. It includes web server log, proxy server log, web browser logs, etc. and the logs are created when users communicate with the web server. The web logs facilitate web administrator to identify the users, their location and browsing pattern etc. It allows identifying website. It also stores the information such as IP address, referring website, timestamp, browser used, used platform etc. The interesting information collected from this web log helps website administrator serving effectively the needs of the users visiting their website efficiently.

Web usage mining focuses on two points- 1) how the website administrators want their website to be used by the user and 2) how the user actually uses this websites. The deviation between the actual and expected use can be reduce after recognizing and personalizing website according the actual need of user. The main focus of this paper is to classify the user in the basis of discover patterns from web usage log which gets created when user interact with the web server.

2. LITERATURE SURVEY

Satya Prakash Singh, Meenu [1] discussed different tools and techniques for web usage mining. They show how Patterns are discovered by making various techniques like statistical analysis, association rule, clustering, classification, sequential pattern. They include of knowledge query mechanism and intelligent agent improvement the efficiency of pattern analysis.

J. Umarani, K. Karpagam [2] they mainly concentrates on methods applied in user identification phase of data preprocessing and prepared analysis of these methodologies to find out the more appropriate on web server log.

K.Dharmarajan , M. A. Dorairangaswamy [3] author describe how to discover patterns from browsing and navigation data of web users. They mainly focus on extraction of user frequent access page using web log data. This complete analysis work has been implemented in the Web Log Expert tool.

Nandita Agrawal, Anand Jawdekar [4] they discussed the key components of web mining and describe various data preprocessing method on the basis of various factors like data, input, output, memory required, working concept, complexity etc.

F. Mary Harin Fernandez, R. Ponnusamy [5]
In this author propose a new web log mining method for determining web access procedure from interpreted web usage logs. In the first step data cleaning and preprocessing method is used for cleaning unwanted data logs and in second step analysis is done of processed data for several user behavioral applications. Third step, the relevant grade and decision making concept are applied to each user's entries in order to shrink users' search period.

Tapan Kumar, Tapesh Kumar [6] author provides the semantic web approach of web usage mining. In this they uses the KNIME tool for the web usage mining.

M. Aldekhail [7] explains in detail the process of web usage mining and the different techniques used for pattern discovery. Also, discussed different applications and tools used for web usage mining. Provides three types of clustering methods such as partitioning, hierarchical and model-based methods for pattern discovery and also describe classification, association rule mining and sequential rule mining method.

Virendra R. Rathod, Govind V. Patel [8] Authors focus on user Behavior using web log file prediction using web log record, click streams record and user information. They proposed Web using mining technique to predict browsing behavior of user may visited based on previous history and knowledge using FCM(Fuzzy C-Means)Algorithm and Markov Model.

Parth Suthar, Prof. Bhavesh Oza [9] they describe various web usage mining techniques, classified based on their nature, which have been developed to find useful information from the Web and also discussed the advantages and drawbacks of these techniques.
M. Santhanakumar, C. Christopher Columbus [10] authors describes analysis of web usage data by applying two different clustering algorithms such as K-means and Fuzzy C means in web usage based dataset using the tool RapidMiner was performed. In this the important steps such as preprocessing, similarity measurement and clustering methods are used for improve the input efficiency, find out user and session similarities and grouping the similar data respectively. The related data are grouped based on the cluster centroid and also the experimental result shows the performances of both clustering algorithms.

Pooja Kherwa, Jyotsna Nigam [11] authors mainly focused on the sequence of tasks of preprocessing of web log files. The complete preprocessing task comprises of multiple phases like field extraction, data cleaning, user identification, session identification, path completion. During this they found that the complete preprocessing strategy is although comprises of multiple phases, but all the phases and technology to implement these phases are interlinked.

Janhavi Bhalerao, Ratna Kendhe, Lahar Mishra [12] author explores the different techniques of web mining with emphasis on web usage mining. They also give description of these methods and their advantages. The distinction between web mining types is also introduced. Association rule mining and clustering algorithms are suggested to produce effective results of discovering web usage patterns.

R. Suganya [13] author presents the important concepts of Web usage mining and its various practical applications. A novel approach called intelligent-miner" (i-Miner) is presented. The i-Miner hybrid framework optimizes a fuzzy clustering algorithm using an evolutionary algorithm. The raw data from the logs are cleaned and pre-processed and a fuzzy C means algorithm is used to identify the number of clusters.

Amit Vishwakarma, Kedar Nath Singh [14] they show different kinds of web usage mining techniques with their basic models and proposed Web log analysis method for analyzing and finding informative patterns which are based on the visual clustering method.

M.Rekha Sundari, Y.Srinivas, PVGD.Prasad Reddy [15] authors describes various pattern discovery techniques of WUM and gives comparison of all these techniques. Association rules are used to discover pages that are visited. Sequential patterns are discovered from web access logs which can be used for predicting future visits of the users.

Anshul Bhargav, Munish Bhargav [16] authors proposed a framework which is based on three steps which are as preprocessing, pattern discovery and users classification. User classification is done
on the basis of three parameters such as country based, site entry based and access time based classification.

Sanjeev Dhawan, Swati Goel [17] focuses on web log file overview, its types and the process of web usage mining. The web log files records information of each click by the user. Preprocessing is done to remove unnecessary data from log file. User and session identification is also done as a preprocessing step which includes identifying users and sessions. Pattern discovery techniques such as association rule mining (such as Apriori), clustering and classification are applied on the reduced log file.

P.Nithya, Dr. P.Sumathi[18] they describe all the phases of web usage mining. They Also discussed the reviews of all the phases of WUM. Various techniques of pattern analysis and pattern discovery are mentioned.

NeetuAnand, Prof (Dr.) SabaHilal [19] author present an approach to identify user access pattern from web log data .In the first phase, the server raw log data is preprocessed. In second phase analysis is performed to identify access pattern of users. And at the last Visualization and interpretation of the discovered and
Interesting patterns has done.

V.Chitraa, Dr. Antony Selvdoss Davamani [20] author mainly focus on preprocessing step of web usage mining and give the reviews on existing work done in the preprocessing stage of web usage mining.

In our work we have to first collect the dataset of web log files entries. After this, the collected entries are given to the preprocessing phase and the cleaning of unwanted data takes place. In the second step clustering method as "Bisecting K- means Clustering Algorithm" is used for grouping the similar data sets. After getting the group of similar clusters, in the next step classification is done. The Support Vector Machine algorithm is used for classification. This will helps for pattern recognition.SVM will take all the training data and then classify the data into the given class according to training data which is provided to it.

3. CONCLUSIONS

Web usage mining helps administrator to serve the needs of their website user. It is based upon the discovery and analysis of web usage pattern from web log. For enhancing the usability of web design, the improvement of customer's relation and improving the requirement of system performance Web Usage Mining play very crucial role. It also provides the support for the Web Site Design, providing personalization server and other business making decision, etc. Classification of Web User can be done on various parameters such as based on country, site, and time, so on. In future work support vector machine (SVM) algorithm used for the user classification. The use SVM algorithm improves the accuracy and enhances the result of the classification.

REFERENCES

[1] Satya Prakash Singh, Meenu "Web Usage Mining Tools &Techniques: A Survey "October-2016 IJARIIT.

[2] J.Umarani, K.Karpagam "Investigation of User Identification Methods in Pre-Processing Phase of Web Usage Mining" Augest-2016 IJESC

[3] K.Dharmarajan , M. A. Dorairangaswamy "Discovering User Pattern Analysis from Web Log Data using Weblog Expert" November 2016 IJST

[4] Nandita Agrawal, Anand Jawdekar "A Survey Report On Current Research And Development of Data Processing In Web Usage Data Mining " 2016 IJDTA

[5] F. Mary Harin Fernandez , R. Ponnusamy "Data Preprocessing and Cleansing in Web Log on Ontology for Enhanced Decision Making " 2016 IJST

[6] Tapan Kumar, Tapesh Kumar "WEB USAGE MINING USING SEMANTIC WEB APPROACH: A STUDY, SURVEY AND ANALYSIS " 2016 ICRISEM

[7] M. Aldekhail "Application and Significance of Web Usage Mining in the 21st Century: A Literature Review" February 2016 IJCTE

[8] Virendra R. Rathod, Govind V. Patel "Prediction of User Behavior using Web log in Web Usage Mining" April 2016 IJCA

[9] Parth Suthar, Prof. Bhavesh Oza "A Survey of Web Usage Mining Techniques" 2015 IJCSIT

[10] M.SANTHANAKUMAR, C.CHRISTOPHER COLUMBUS "Web Usage Based Analysis of Web Pages Using RapidMiner "2015

[11] Pooja Kherwa, Jyotsna Nigam "Data Preprocessing: A Milestone of Web Usage Mining "2015 IJESIT

[12] Janhavi Bhalerao, Ratna Kendhe, Lahar Mishra "Techniques for Web Usage Mining" October 2015 IJSRP

[13] R. Suganya "Analysis of Implementating Web Usage Mining and I-Miner in E-Business" 2015 IJARCSSE

[14] Amit Vishwakarma, Kedar Nath Singh "A Survey on Web Log Mining Pattern Discovery "2014 IJCSIT

[15] M.Rekha Sundari, Y.Srinivas, PVGD.Prasad Reddy "A Review on Pattern Discovery Techniques of Web Usage Mining"2014 IJERA

[16] Anshul Bhargav, Munish Bhargav "Pattern Discovery and Users Classification Through Web Usage Mining" 2014 ICCICCT

[17] Sanjeev Dhawan, Swati Goel "Web Usage Mining: Finding Usage Patterns from Web Logs" 2013 AIJRSTEM

[18] P.Nithya, Dr.P.Sumathi "A Survey on Web Usage Mining: Theory and Applications" August 2012 IJCTA

[19] NeetuAnand, Prof (Dr.)SabaHilal "Identifying the User Access Pattern in Web Log Data" 2012 IJCSIT

[20] V.Chitraa Dr. Antony Selvdoss Davamani "A Survey on Preprocessing Methods for Web Usage" Data 2010 IJCSIS