DIFFERENT WEB USAGE MINING TECHNIQUES TO PRODUCE EFFECTIVE PATTERN ANALYSIS RESULTS FOR WEB USAGE DATA

[1] Anshu Dixit, [2] Dr. Pritaj Yadav

[1]Student, [2] Asso.Prof

[1]Computer Science Engineering, RNTU, Bhopal. M.P.

Abstract: Effective pattern analysis, we use web usage mining process to extract useful information from web servers that is browsed and accessed. This shows the browsing and access patterns of the users and it is based on user's needs, web usage technique is used to discover hidden patterns access from weblog with the help of different types of mining algorithms for help effective analysis of pattern. This process is guided by three phases: preprocessing of data, pattern discovery, and analysis. Now present in each cycle in detail, is the process of extracting useful information from data files that shows effective analysis of the specific set of patterns for different web data sets and effective pattern analysis.

Index Terms - **Web Usage Mining, Classify, Cluster, Associate rules.**

I. INTRODUCTION

Every day we have to go through several kinds of information that we need, and what do we do? Just browse the web and the desired information is with us on a single click. Today, the internet plays such a vital role in our everyday lives that it is very difficult to survive without it. The World Wide Web (WWW) has influenced a lot of both users (visitors) and web site owners. The website owners are able to reach all the targeted audiences nationally and internationally. They are open to their customers 24x7. On the other hand, visitors are also availing those facilities.

In the last 15 years, the growth in the number of websites and the number of visitors to those websites has increased exponentially. Due to this growth, a huge quantity of web data has been generated. Data mining techniques can be applied to mine the interesting data from this huge pool. But the web data is unstructured or semi-structured. So we cannot apply the data mining techniques directly. Web usage mining is the process of extracting useful information from datasets, e.g., uses Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web based applications.

II. PROPOSED ALGORITHM

2.1 Apriori Algorithm –

To improve the efficiency of level-wise generation of frequent item sets, an important property called the Apriori property is used, which helps by reducing the search space. The Apriori algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule leaning that analyzes whether people who bought product A also bought product B.

The Apriori algorithm finds the frequent sets L in Database D.

Find a frequent set, $L_{k\,1}$.

Join Step.

C_k is generated by joining $Lk\,1$.

Step.

Any $(k\,1)$ item set that is not frequent cannot be a subset of a frequent k- itemset, hence should be removed.

where (C_k: Candidate item set of size k)

(L_k: frequent itemset of size k)

2.2. Dynamic Itemset Counting (DIC) Algorithm-

It is an alternate to Apriori Item set generation in which Itemsets are dynamically added or deleted as transactions are read. It relies on the fact that for an itemset to be frequent, all of its subsets must also be frequent, so we only examine those itemsets whose subsets are all frequent. The algorithm stops after every M transaction to add more itemsets.

2.3 Clustering

The main purpose of clustering in web usage mining is to aggregate similar sessions together. Self organized maps, graph partitioning, ant based technique, K-means with genetic algorithms; EM-C Fuzzy means algorithms are the algorithms used for clustering the sessions. Mainly, clustering is two types: usage clustering and page clustering. Usage clusters discover users that have some browsing pattern where page clusters gather the content related pages together. This is pre-formed by transforming the session into vectors of N elements, where N may be no of pages or page views. Clustering helps in personalizing the web site as it identifies the users with similar behavior.

2.4 The K-Means Clustering method

It is the simplest and most popular classical clustering method. The method is called the K-Means since each of the K clusters is represented by the mean of the objects (called the centroid) within it. It is also called the centroid method since at each step the centroid point of each cluster whose centroid is assumed to be known and each of the remaining points are allocated to the cluster whose centroid is closest to it. Once this allocation is completed, the centroids of the clusters are recomputed using simple means and the process of allocating point to each cluster is repeated until there is no change in the clusters.

2.5 Classification

Classification is supervised way of learning which maps the data items to one of the many predefined classes. Various supervised learning algorithms used for doing classification are decision tree, naïve Bayesian classifiers, k-nearest neighbor classifiers and support vector machine. Classification mainly performs the automatic categorization of documents. In web usage mining, application of classification algorithms on server logs may lead to detection of interesting pattern such as 40% of users who visits news site are in the age group of 30-35 years.

III. EXPERIMENT AND RESULT

A collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions. it is now used in many different application areas, Its use any modern computing platform. Comprehensive collection of the data preprocessing and modeling techniques. Ease of use due to its graphical user interfaces.

- All of use techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes. The Preprocess panel has facilities for importing data from a database, a comma separated values file, etc., and for preprocessing this data using filtering. These filters can be used to transform the data and make it possible to delete instances and attributes according to specific criteria.
- The Classify panel enables applying classification and regression algorithms to the resulting dataset, estimating the accuracy of the resulting predictive model, and visualizing erroneous predictions, or the model itself .
- The Associate panel provides access to association rule learners that attempt to identify all important interrelationships between attributes in the data.
- The Cluster panel gives access to the clustering techniques' applications, e.g., the simple k-means There is also an implementation of the expectation maximization algorithm for learning a mixture of normal distributions.
- The Select attribute panel provides algorithms for identifying the most predictive attributes in a dataset.

- The visualize panel shows a scatter plot matrix, where individual scatter plots can be selected, enlarged, and analyzed further using various selection operators.
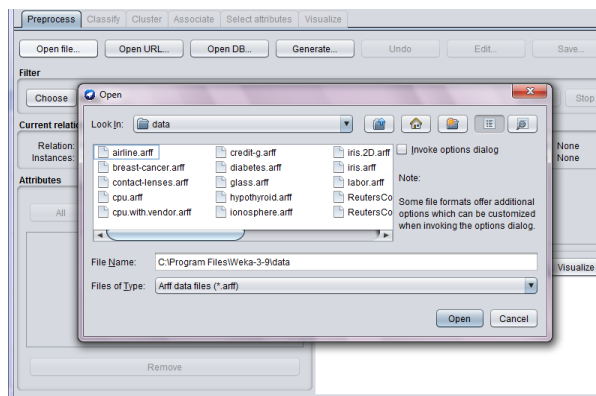

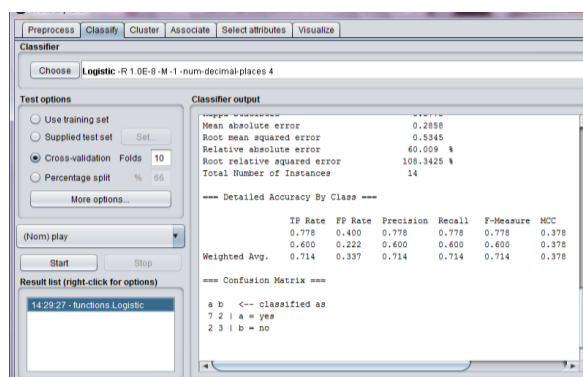Figure 1: Preprocessing the open data set
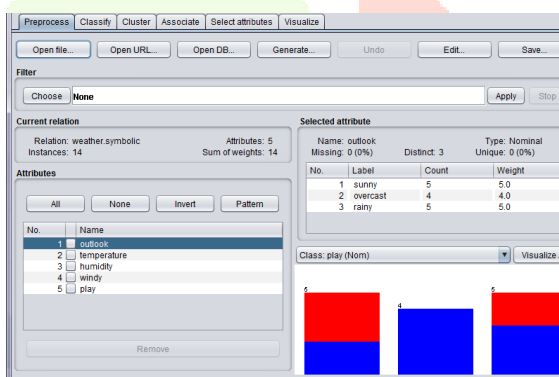

Figure 2: Preprocessing Data

Figure 3:Analysis Explorer

IV. CONCLUSION

The key computer science contributions made by the field, a number of prominent applications, and outlined some areas of future research. Our hope is that this overview provides a starting point for fruitful discussion. Data mining is the study of exploring patterns in huge volumes of raw data. The term Web mining has been used to refer to techniques that help us to find content of web and retrieve the user's interest and needs. This paper focuses on the comparison of mining algorithms. Web Usage Mining and various criteria are considered for pattern discovery. As huge data is adding in repository every second so there is need of quality information to satisfy upcoming needs of user. In future Web Service provisioning and web service discovery should be made by analyzing user interest.

This work explores the different techniques of web mining with emphasis on web usage mining. A detailed description of these methods and their advantages is given. The distinction between web mining types is also introduced .Overall the usage mining process is illustrated. Various combination of algorithms like association rule mining and clustering are suggested to produce effective results of discovering web usage patterns. Since the web continues to develop in size and complexity with time, it has become difficult to find information which is to the point. Hence, web data mining techniques are developed to get useful data from the web page. Web log pre-processing is initial necessary step to improve the quality and efficiency of the later steps of web usage mining.

REFERENCES

[1] "Performance Evaluation of the MapReduce-based Parallel Data Preprocessing Algorithm in Web Usage Mining with Robot Detection Approaches", IETE 2021.

[2] Akash Yadav1, Minal Chudasama2, "Unauthorized Terror Attack Tracking System using Web Usage Mining", International Research Journal of Engineering and Technology, 2020.

[3] Rajinder Singh Rao1, Jyoti Arora2, "A Survey on Methods used in Web Usage Mining", International Research Journal of Engineering and Technology, 2017.pp 2627-2631.

[4] M. Aldekhail, "Application and Significance of Web Usage Mining in the 21st Century", International Journal of Computer Theory and Engineering,2016.

[5] Parth Suthar, Prof. Bhavesh Oza, "A Survey of Web Usage Mining Techniques" International Journal of Computer Science and Information Technologies, pp 5073-5076, 2015.

[6] Mitali Srivastava, Rakhi Garg, P. K. Mishra, "Preprocessing Techniques in Web Usage Mining: A Survey", International Journal of Computer Applications,2014.

[7] A. A. Abd El-Aziz 1, P. Senthil Pandian2, "A Framework for Clustering & Enhanced Approach for Frequent Patterns in Web Usage Mining", International Journal of Advanced Science and Technology, 2020.

[8] Pranit Bari,, P.M. Chawan "Web Usage Mining", Journal of Engineering, Computers & Applied Sciences,2013.

[9] D. Uma Maheswari, Dr. A. Marimuthu, "A Study of Web Usage Mining Applications and its Future Trends" International Journal of Engineering Research & Technology,2013.

[10] Vijayashri Losarwar, Dr. Madhuri Joshi, "Data Preprocessing in Web Usage Mining", International Conference on Artificial Intelligence and Embedded Systems, 2012.

[11] V.Chitraa, Dr.Antony Selvadoss Thanamani, " A Novel Technique for Sessions Identification in Web Usage Mining Preprocessing, International Journal of Computer Applications,2011

[12] A. Anitha, "A New Web Usage Mining Approach for Next Page Access Prediction", International Journal of Computer Applications,2010