ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022

A STUDY OF WEB USAGE MINING: FINDING PATTERNS IN WEB LOG DATA

Manish Patankar[1], Gururaj Dangare[2], Priya Mathurkar[3]
Pratibha Institute of Business Management, Chinchwad, Pune, India

ABSTRACT:

Web mining is the application of data mining techniques to discover patterns from the World Wide Web. Personal data of an individual on web is not only data related to personal profile but also is related to the navigation or surfing that individual do on web by using search engines. Web log is a listing of page referenced data. The data includes stuff data, face book details, warranty card, reward card, information relate to what an individual bye frequently, to the website an individual visits frequently, shop frequently etc. The sequence of clicked data can be used to know the users. All these data is stored as web log data. This data needs to be processed and analyzed to use it to know the interest of a visitor or customer to increase business. This paper gives a brief study of how a web log data is preprocessed and how patterns are formed to detect the most likely user and group of users traversing certain web pages.

Keywords: Web mining, web logs, preprocessing, path completion, pattern discovery, pattern analysis

INTRODUCTION

Web mining is the application of data mining techniques to discover patterns from the World Wide Web. This pattern may be the data that is actually present in web log or data related to web activity. Web log data is a website that consists of series of entries arranged in reversed chorological order, often updated on frequently with new information about particular topics [13]. Web mining is the area of data mining. It consists of majorly three subareas: (i) Web content mining, (ii) Web structure mining and (iii) Web usage mining. Fig 1.1 below shows Web mining taxonomy.

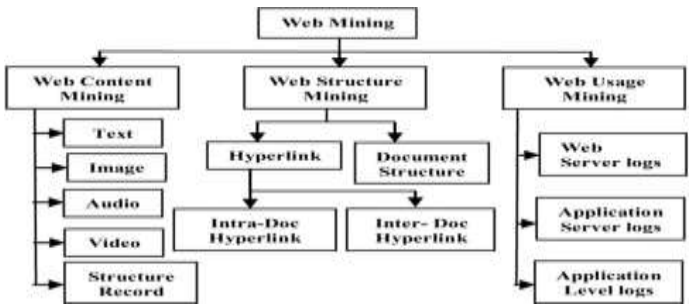

Fig. 1.1 Web Mining Taxonomy

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

Web Usage Mining can be thought as extending the work performed by basic search engines [4]. Web log data provide information about activities performed by users from the moment the same user leave it. This record of user's action within a website are stored in a log file [14]. To find common patterns and interest of the user's, web log data is to be preprocessed and mined.

RECENT WORK

Many related study and literature reviewed on this topic has been done earlier. Satpal Singh, Vivek Badhe have suggested the different view point for finding the web-user on the basic of temporal approach. Such kind of analysis could be useful for target marketing based on time or for web services optimization in 2014 [8]. Rachit Adhvary focuses on the concept of Web Usage Mining, specifies the importance of web usage mining, brief details on the Pattern Matching and other different functionalities of data mining used in web usage mining in 2013[9]. M. Aldekhail illustrates the different applications and tools used for web usage mining. Finally, it explains some current issues and challenges such as privacy and scalability, which are important issues in web usage mining in 2016[10].

METHOD

Web usage mining uses the web log files which are resided at web servers, proxy servers and browser machines as a source to identify user's website access behaviors. The users website visiting details are recorded in various sources in common log format. The web logs are massive in size and not lying in appropriate format. Fig.3.1 below a typical web log data format.

Fig.3.1. Web log data[13].

Web usage mining actually consists of three separate types of activities:

a. Preprocessing
b. Pattern Discovery

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

c. Pattern Analysis

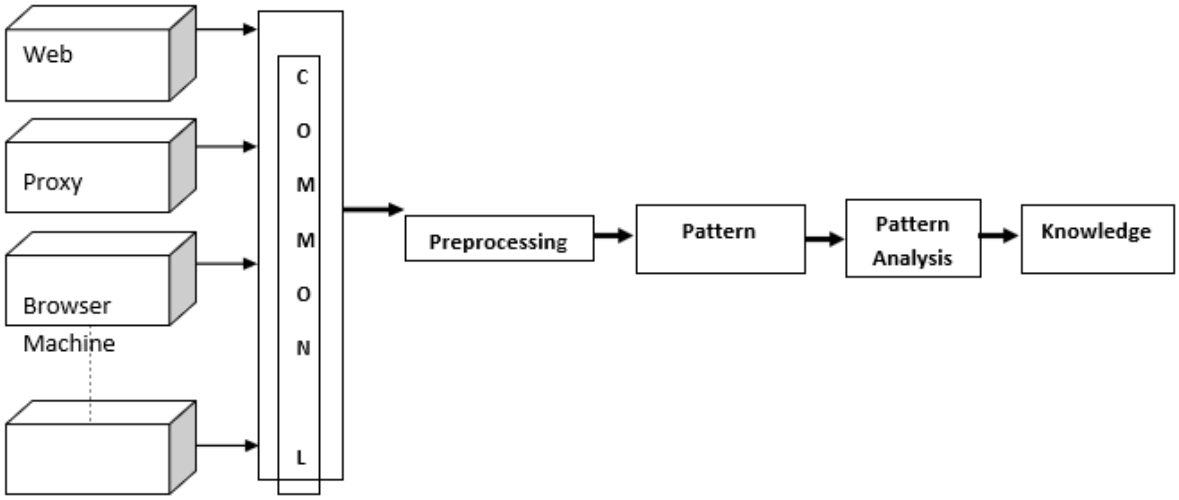

Fig. 3.2 above shows collection of web-log data and pattern recognition. Web log data is collected from various sources are preprocessed to discover interesting usage patterns from web data. So, careful preprocessing is applied to make the web logs suitable for extracting knowledge. Pattern analysis techniques are applied to the preprocessed web logs to obtain the information from them [6]. Knowledge discovered is then understand in order to better serve the needs of web-based applications to improve customer satisfaction, digital sale conversion, marketing results, branding and improved website metrics as well as for advertising.

a. Preprocessing:

The data is probably not in format that is usable by mining applications. When any data is to be used in a mining application the data may need to be reformatted and cleansed. Preprocessing includes following steps:

1. Cleansing and reformatting – Before using data for mining ,data has to reformatted and cleansed
2. User identification- user can be identified through user ID and IP address.
3. Session identification- user visit same page number of times during one logical time. This shows the interest of user for the same page.
4. Path completion –It checks the last page referred and,track and counts missing pages.
5. Path completion- It is atechnique to add pages which are accessed but do not exists in logs.

To keep track of patterns identified during mining process, a data structure name trie-a digital tree has been proposed is used. It is used to store string for pattern-matching applications. Fig.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

3.2 Data structure Trie and Compressed Trie for searching patterns

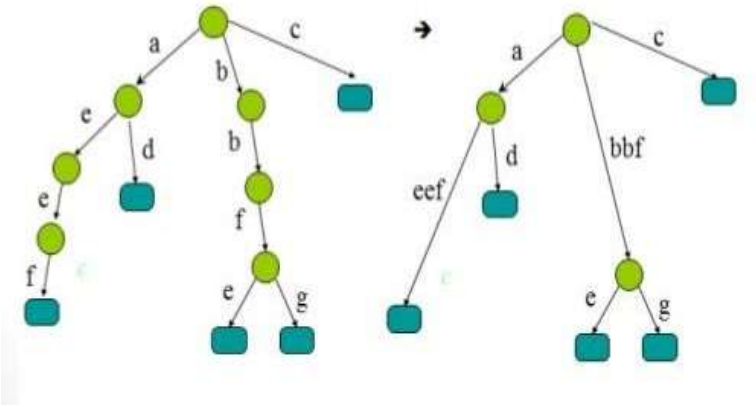

Fig. 3.2 Data structure Trie and Compressed Trie for searching patterns [13].

Trie help to find subsequence in a sequence of web pagesand also finds common sequences among multiple sequences.

b. Pattern Discovery:

Different traversal path are used to traverse user session. The web log patterns are divided into several clusters called sub logs for similar interest. The steps for searching the pattern may be repeatedly applied till the "most likely" searching patterns are found. There are several pattern discovery algorithms. There are several clustering methods that can be deliberately used to discover patterns.

c. Pattern Analysis:

Some of the patterns may be of interest and some may be non-informative. Analysis of the web log data can include simple query and reporting functions. Various pattern analysis methods can be used like statistical analysis, comparing traversal paths of the users, filtering traversal path with a desirable pattern, tracing out the sequences with wild card, running queries, coloring common patters. Some of the techniques are:

- Queries and Reports(written in SQL)
- Managed query Environment(Centralized control of reporting)
- OLAP and OLAP variant, etc.

CONCLUSION

This paper gives a brief study of different activities of web usage mining. Searching and analysis of web data has become today's need since data is growing day by day. To better understand the use of web and to get knowledge about users in-depth study of the activities of

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, S Iss 2, 2022**

user using web is needed. Such kind of study may help to explore the business and attract the customers or users for the business.

REFERENCES

1. Athena Vakali1, Jaroslav Pokorny2, and Theodore Dalamagas,"An Overview of Web Data Clustering Practices",2013.
2. Ananthi.J," A Survey Web Content Mining Methods and Applications for Information Extraction from Online Shopping Sites", International Journal of Computer Science and Information Technologies,2014.
3. SupinderSingh,SukhpreetKaur,"Web Log File Data Clustering Using K-Means and Decision Tree", International Journal of Advanced Research in Computer Science and Software Engineering,August 2013.
4. HemanshuRana, Mayank Patel, "A Study of Web Log Analysis Using Clustering Techniques", International Journal of Innovative Research in Computer and Communication Engineering, June 2013
5. PoojaMehtaa, Brinda Parekh, KiritModi, and PareshSolanki," Web Personalization Using Web Mining: Concept and Research Issue", International Journal of Information and Education Technology, October 2012.
6. R. Suguna, D. Sharmila, "clustering web-log files-A review", International Journal of Engineering Research & Technology, April 2013.
7. SatpalSingh1 ,VivekBadhe," An Exclusive Survey on Web Usage Mining For User Identification", International Journal of Innovative Research in Computer and Communication Engineering 2014.
8. Rachit Adhvaryu ,b. h. GardiVidyapith, "A Review Paper on Web Usage Mining and Pattern Discovery", Journal of Information, knowledge and Research in Computer Engineering, Nov 12 to Oct 13.
9. M. Aldekhail," Application and Significance of Web Usage Mining in the 21st Century: A Literature Review", International Journal of Computer Theory and Engineering,February 2016.
10. Mr. Akshay Upadhyay, Mr. BalramPurswani," Web Usage Mining has Pattern Discovery", International Journal of Scientific and Research Publications, February 2013.
11. Richa Patel,Akshay Kansara,""Web Usage Mining- A survey on User's Navigation pattern from Web Logs", International Journal of Scientific Research and Development 2014.
12. www.wikipedia.com.
13. Margaret H.Dunham, "Data Mining Introductory and Advanced Topics".