Current Trends And Future Challenges In Web Usage Mining

Mr.J.Jelsteen [1], Rohitt Abiman S[2], Sankareshwaran S[3]
[1]Asst.Professor, Dept of BCA
[2, 3]Dept of BCA
[1, 2] Affiliated to Bharathiar University, Coimbatore,India

Abstract- *The collection, analyzation and summarization of web activities from web users is called as web usage mining. In current trend web usage mining has no high quality and sufficient security. Retrieving the required web page from the web efficiently and effectively becomes a challenging task because web is made up of unstructured data, which delivers the large amount of information and increase the complexity of dealing information from different web service providers. The collection of information becomes very hard to find, extract, filter or evaluate the relevant information for the users. There are many flaws in current techniques which will be a challenge in future web usage mining. In this paper we have studied the concepts of web usage mining, its current trends and its future challenges from previous researches.*

Keywords- Data mining, web mining, web usage mining, current trends, limitations, future challenges.

I. INTRODUCTION

Data mining is a process by which large amount of raw data is filtered into a useful information to look for patterns in large clusters. Millions of terabytes of data are getting generated and companies are processing it using big data technologies. Look at the industry you are working, your consumer behaviour - what they buy, what they liked, which location they are currently in, their age and a lot of other factors influences buying behaviour. All these data points are to be collected.

Another point of view is that we require data from a website. Only top service providers can ensure your data gets treated correctly. Go for someone who can assure you top notch Security, track records and foremost knowledge of processing similar data with domain knowledge. It is used in market analysis and management, corporate analysis and risk management, fraud detection, Ecommerce product data mining, E-commerce web scraping, Price monitoring and the list goes on.

Web mining is to discover and extract information from documents and web services. It is used to extract knowledge. It is categorized into three steps such as web content mining, web structure mining, web usage mining.

It is used in several fields like e-commerce, business intelligence, web services, system performance, knowledge management. Web content mining is the process of filtering useful information from the content of the user's document. Web structure mining is the process of getting structural information from the web. The large dataset of transactions, which are previous users' web browsing information, are searched for those web pages that occur in a frequent manner.

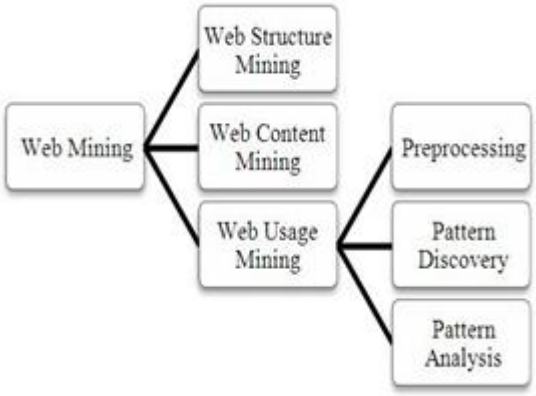
II. WEB USAGE MINING

Web usage mining is the discovery of web activities done by the user. It is based on statistical methods. Extraction technique is mainly used to mine the data. This invisible data gathering is common on the web. Knowledge collected through mining the web could be a threat to people, when, for instance, personal data is misused, or is used for a purpose other than the one for which it is supplied. This same knowledge, however, can bring lots of advantages. Knowledge discovered through data mining is important for all categories of applications involving planning and control. There are some specific benefits of web-data mining like improving the intelligence and efficiency of search engines. Web-data mining can also be used for marketing intelligence and experience by analysing a web user's online behaviour and turning this information into marketing information.

SUBTASKS IN WEB USAGE MINING

1. Resource finding: the task of retrieving the Web documents.
2. Information selection and pre-processing: automatically selecting and pre-processing specific information from retrieved Web resources.
3. Generalization: automatically discovers general patterns at individual web sites as well as across multiple sites.
4. Analysis: evaluation or interpretation of the mined patterns.

The categories of web usage mining are giving in the following diagram.



DATA PREPROCESSING

Data pre-processing is the combination of four steps mainly data cleaning, session identification and reconstruction, content and structure retrieving, and data formatting.

PATTERN DISCOVERY

The pattern discovery techniques in web usage mining are statistical analysis, association rules, clustering, classification and sequential patterns.

PATTERN ANALYSIS

The final step in web usage mining is pattern analysis. During this analysis fascinating and non-fascinating patterns from the pattern discovery unit is separated.

APPLICATIONS

Applications of web usage mining are system performance improvement, personalization, site modification, e-commerce and so on.

PERSONALIZATION

Web personalization is the process of customizing a Web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user's behaviour in correlation with other information collected in the Web context. Due to the explosive growth of the Web, the domain of Web personalization has gained great momentum both in the research and commercial areas.

E-COMMERECE

Various organizations and companies are also employing the web in order to introduce their products or services around the world. So E-commerce or electronic commerce is formed. E-commerce is any type of business or commercial transaction that involves the transfer of information across the internet. So a huge amount of information is generated and stored in the web services. This information overhead leads to difficulty in finding relevant and useful knowledge, therefore web mining is used as a tool to discover and extract the knowledge from the web. This massive increase in the uptake of e-commerce has led to a new generation of associated security threats. We use web mining techniques for security purposes, in detecting, preventing and predicting cyber-attacks.

SUPPORT TO DESIGN

Usability is one of the major issues in the design and implementation of web sites. The results produced by Web Usage Mining techniques can provide guidelines for improving the design of web applications. Adaptive Web sites represents a further step. In this case, the content and the structure of the web site can be dynamically reorganized according to the data mined from the users' behaviour.

ALGORITHMS

Apriori algorithm is the most popular algorithm that is used to find frequent item sets from an enormous database of transactions.

SEQUENTIAL PATTERNS

Sequential patterns are used to discover the subsequence in the large volume of sequential data. In web usage mining, sequential patterns are used to find user navigation patterns which appear frequently at meetings. A sequential pattern is often as follows: 70% of users who have first observed the page A1 website and then page A2 website observed page A3 website the same session, too. The sequential patterns may seem to association rules. Actually, algorithms that are used to extract association rules, can also be used to generate sequential patterns. But the sequential patterns are included the time, it means that the sequence of events occurred is defined in

sequential patterns. In the above example, pages A, B, C are respectively seen in a user session. Two types of algorithms are used for mining sequential patterns. The first type of algorithms is based on association rules mining. In fact, many common algorithms of mining sequential patterns have been changed for mining association rules. The second

type of sequential patterns mining algorithms have been introduced in which the tree structure. Evaluation results show that its performance is higher than an algorithm such as GSP.

CLUSTERING

Clustering techniques contains groups of similar items among large volumes of data. This is done based on distance functions which measures the degree of similarity between various items. Clustering in web usage mining is used for grouping similar contents. Two types of interesting clustering can be found in this area: user clustering, page clustering. Clustering of user details is usually used to analyse the tasks in web mining and web analytics. More knowledge derived from clustering is used to divide the market in e-commerce. Using the similar content graph and the amount of time spent viewing a page to estimate the similarity of contents. In clustering technology, first the repetitive patterns are obtained from the user's sessions by using association rules. Then, these patterns used to construct a graph where the nodes are the visited pages. The edges of the graph connect two or more pages, if these pages exist in a pattern obtained so the weight will be given to the edges that shows the relationship between the nodes. Then, for clustering, the graph is recursively divided to user behaviour groups are detected. The main objective of web usage mining is to collect data about the user's moving patterns. This information can be used to improve the web sites in the user view. Three main applications of this mining are studied in this section. The privatization of web content Web usage mining techniques can be used for personalization of web users. For example, user behaviour can be instantly predicted by comparing the user's current survey patterns with survey patterns extracted from the log files. Recommendation systems which have a real application in this area are, suggest links that direct the user to his favouritepages.

CURRENT TRENDS

Web usage mining is used to identify internet fraud, cybercrime, illegal online gambling, terrorism, hacking, virus spreading and so on. It is used as an analysing tool. Current techniques of web usage mining support keywords, links, content, secondary data etc... Nowadays, web usage mining has no enough security and causes damage to the personal data of web users. Web usage mining plays an essential role in this world of modern technology. The current web is based on HTML (hypertext mark-up language).

LIMITATIONS

Web usage mining by its own does not create any issues, but this methodology when used on personal data might cause issues. The most criticized ethical issue involving web usage mining is the indulgence of private data. Privacy is considered lost when data containing an individual is obtained, used, or categorized, especially if this occurs without their knowledge.

The grouped data will be analysed, and grouped to form profile data; the data will be made unknown before clustering so that there are no personal profiles. Thus these applications take the individuality of the users by judging them by their mouse clicks. De-individualization, can be defined as a process of judging and treating people on the basis of group attributes instead of on their own individual attributes and merits. Another important concern is that the companies collecting the data for a specific purpose might use the data for totally different purposes, and this essentially violates the user's interests.

THREATS

The growing trend of selling personal data as a commodity encourages website owners to trade personal data obtained from their site. This trend has increased the amount of data being captured and traded increasing the likeliness of one's privacy being invaded. The companies which buy the data are obliged make it anonymous and these companies are considered authors of any specific release of mining patterns. They are legally responsible for the contents of the release; any inaccuracies in the release will result in serious lawsuits,butthereisnolawpreventingthem from trading the data. To us it might look like there is no real threat but our personal data could be stolen.

There is another type of web usage mining that is the use of forms on the web. Web users often have to fill out a form to simply gain access to a web site. Or, there are fields to be filled in on online ordering forms, which are of no importance to the purchase. When a user fills in an online form of any kind, the data the person shares can be used for customer profiling. By sending back the form, the web log also registers the current IP address of the web user, and person's personal data can, therefore, be linked to person's browsing behaviour on that particular web site. Although a user decides whether to use a particular website or not their personal data has been exploited.

FUTURE CHALLLENGES

The future goal of web usage mining is to find and retrieve information in a safe manner. The concept of collaborative filtering should improve in future. Web usage

mining in various fields like medicine, e-learning, e-commerce (understanding the needs of customers), offline data security should be developed in future.

The problem of machines not able to understand the web content should be changed by high web technologies for better work in web usage mining. The limitations such as threat during web usage mining and loss of personal data of web users must be prevented. The future work of web usage mining is to use the algorithm in the techniques of web mining by analysing the web user's using pattern. Businesses which have been slow in accepting the process of data mining are now catching up with the others. Extracting important information through the process of data mining is broadly used to make critical business decisions. In the coming period, we can think data mining to become as ubiquitous as some of the more dominant technologies used today. Some of the key data mining trends for the future include like Multimedia processing, Ubiquitous computing, Distributed middleware and Spatial and Geographical Mining.

MULTIMEDIA PROCESSING

Multimedia has at its very core the field of signal-processing technology. With the exploding growth of the Internet, the field of multimedia processing in communications is becoming more and more exciting. Although multimedia leverages numerous disciplines, signal processing is the most relevant. Some of the basic concepts, such as spectral analysis, sampling theory and partial differential equations, have become the fundamental building blocks for numerous applications and, subsequently, have been applied in such diverse areas as transform coding, display technology and neural networks. The diverse signal-processing algorithms, concepts and applications are interconnected and, in numerous instances, appear in various reincarnated forms. We canalso implement it in web usage mining.

UBIQUITOUS COMPUTING

Ubiquitous computing is a process in which the processing of information is linked with each activity or object as encountered. Ubiquitous computing focuses on learning by removing the complexity of computing and increases efficiency while using computing for different daily activities. Ubiquitous computing is also known as pervasive computing, available anywhere and everywhere.

SPATIAL DATA MINING

Spatial data mining is the application of data mining to spatial models. In spatial data mining, analysts use geographical or spatial information to produce business intelligence or other results.

One fascinating thing about the term "spatial data mining" is that it is generally used to talk about finding useful and non-trivial patterns in web data. The core goal of a spatial data mining project is to distinguish the information in order to build real, actionable patterns to present, excluding things like statistical coincidence, randomized spatial modelling or irrelevant results. One-way analysts may do this is by combing through data looking for "same-object" or "object-equivalent" models to provide accurate comparisons of different geographical locations.

V. CONCLUSION

To keep up with technical changes, ongoing debates about ethical issues are an essential part of this combined solution-package and can help prevent possible future harms. As the World Wide Web becomes the most important part of modern society, organisations of all kinds are engaged in efforts to use web-data mining technology for varied purposes. Many of those purposes are of a commercial nature as there is money to be made in the collection and the intelligent analysis of information about people. While most of the benefits go to the web miners, the web users are facing the dangers of web-data mining. Web data is growing at a rapid rate.

Web mining is a florishable area of research with many successful applications. As the Web and its usage continues to perish, so perishes the opportunity to analyse Web data and extract useful information from it. To extract the needed data from web warehouse, the three categories (Web Content Mining, Web Structure Mining and Web Usage Mining) of web mining plays an inevitable role. Web mining is one of the most important applications of data mining. It is having its own benefits and successful applications with which we can overcome the problems or difficulties faced in data mining. As everything has its pros and cons web usage mining also has its own disadvantages.

As ethical issues will grow as rapidly as the technology, ethical considerations should be an integrated and essential part of this development process instead of something at its side. It is probably not possible to develop comprehensive ethical guidelines covering every possible misuse. This is all the more reason to realise the seriousness of the dangers, and to continuously discuss these ethical issues.

This is a joint responsibility of both web miners and web users. In this paper the reason why, web mining techniques have to be improved is clearly depicted.

REFERENCES

[1] A Review on Emerging Trends of Web Mining and It's Applications | ISSN: 2321-9939 IJEDR1303021 INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH | IJEDR Website: www.ijedr.org | Email ID: editor@ijedr.org 98 A Review on Emerging TrendsofWebMiningandIt'sApplications1Mr. Dushyant B. Rathod, 2Dr. Samrat Khanna

[2] ManojPandia, Subhendu Kumar Pani, Sanjay Kumar Padhi, LingarajPanigrahy, Ramakrishna, ―A Review of Trends in Research on Web Mining‖, A Review of Trends in Research on Web Mining International Journal of Instrumentation, Control & Automation (IJICA), Volume 1, Issue 1, 2011.

[3] S. Geetharani, S. PriyadharshiniIJETST-Vol.||02||Issue||03||Pages 1973- 1975||March||ISSN 2348-9480

[4] ANITHA TALAKOKKULA Computer Engineering and Intelligent Systems ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.6, No.2, 2015

[5] M. Aldekhail, International Journal of Computer Theory and Engineering, Vol. 8, No. 1, February 2016

[6] Mitali Srivastava, Rakhi Garg, P. K. Mishra InternationalJournalofComputerApplications(0975 – 8887) Volume 97– No.18, July 2014

[7] Jaydeep Srivastava, Robert Cooley, Mukund Deshpande, and Pang Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data (2000). SIGKDD Explorations, Vol. 1, Issue 2, 2000.

[8] Margaret H. Dunham, Data Mining Introductory and Advanced Topics, Prentice Hall, 2003.

[9] A.G. Buchner, M.D. Mulvenna, Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining, ACM SIGMOD, Vol. 27, No. 4, pp. 54-61, 1998.

[10] BamshadMobasher, Robert Cooley, Jaydeep Srivastava, Creating Adaptive Web Sites Through Usage- Based Clustering of URLs, in Proceedings of the 1999 IEEE Knowledge and Data Engineering sExchange Workshop (KDEX'99), November 1999.

[11] Brogdvision.http://www.broadvision.com.

[12] Like minds.http://www.andromedia.com.

[13] BamshadMobasher, Robert Cooley, and Jaydeep Sri