RESEARCH ARTICLE

An Analysis of Recent Trends and Challenges in Web Usage Mining Applications

C. Sakthipriya[1], G. Srinaganya[2], Dr. J. G. R. Sathiaseelan[3]

[1] Dept. of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India
[2] Dept. of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India
[3] Head, Dept. of Computer Science, Bishop Heber College, Tiruchirappalli, Tamilnadu, India
[1] c.sakthipriyamca@gmail.com; [2] srinaganyag1978@gmail.com; [3] jgrsathiaseelan@gmail.com

Abstract— The World Wide Web is an interactive and popular way to transfer the information. A massive quantity of information is available over the internet. Now-a-days web mining is one of the significant topic in data mining, which is used to extract information from web documents. It is basically categorized into three types, namely, web content mining, web structure mining and web usage mining. It is used to create many web applications, which are playing an important role in our daily life. This paper is mainly aimed to analyze the web mining categories and its web applications.

Keywords— Web mining, Web content mining, Web structure mining, Web usage mining, Web applications of web mining

I. INTRODUCTION

Data mining techniques are applied to web data that refers web data mining or web mining. Web mining is one of the important technique in data mining. It is used to extract useful data from the web that has a hefty amount of documents. It has been explored with large degree and different techniques, which has been proposed for a variety of applications like web search, query classification and personalization [1]. The textual parts of web data consist unstructured data such as free texts, semi-structured data similar to HTML documents, and more structured data - data in the tables or database generated by HTML pages. Web mining copes with semi structured data and unstructured data. It is the most challenging tasks for data mining and data management scholars because there are heterogeneous, less structured data available on the web and easily overwhelm with data. Basically web mining could be divided into the following four step process - Resource finding and Retrieving, Information selection and pre-processing, Patterns and recognition, and Validation and interpretation [2]. The process is symbolized in the Fig. 1.

Resource finding is used to locate unfamiliar documents and services on the web. The process of resource finding is, to extract the data from the web either from online or offline by the user. Information selection and pre-processing are a process, that can select and pre-

process the information automatically, which is retrieved from the web resources. It can be classified into five types they are, data cleaning, user identification, user session identification, access path supplement and transaction identification. Data cleaning is the process to remove the unwanted data which is improving the scope of the data in the web document. User identification identifies the users individually from the web log server. User session identification process is used to recognize the user access information from the web server. Access path estimation is a method to estimate the user access log files from the web server. Transaction identification is a technique, which is based on the user session identification method. Pattern discovery has automatically discovered the patterns from a website as well as across multiple websites [3]. It is divided into five types specifically, path analysis, association rule, sequential pattern, classification and clustering. Path Analysis is a graphical form of any websites. Association Rule is mainly focused to discover the relationship between the web pages and visited web pages by the users of the websites. Sequential Pattern discovery is a technique to find the inner-session pattern of web pages. Classification is a mapping method of data that could be one or several predefined data. Clustering Analysis is used to group the data or items which have similar attributes or characteristics [4]. Pattern Analysis is a process that involves validation and interpretation of mined patterns from the web. The classification of pattern Analysis includes visualization techniques, data and knowledge query, OLAP techniques and usability analysis. Visualization Technique is a method that used to understanding the behavior of web users. Data and Knowledge query is a process that analysis the users' problem or users' needs on the web. OLAP Technique is analysis the process of web servers. Usability Analysis is the analyst software usability and user usability of the websites by accessing any model of the access behavior of the users.
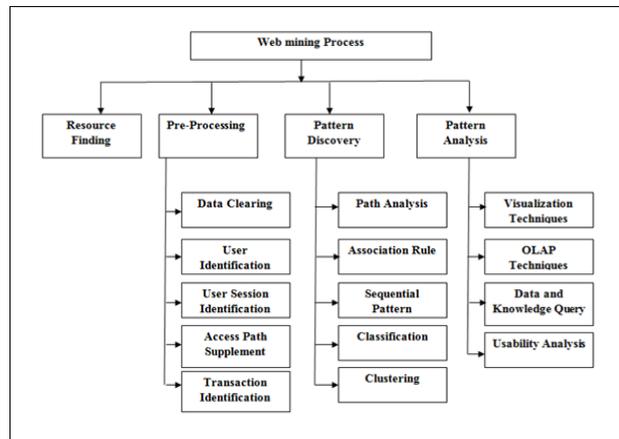

Fig. 1. Web Mining Process

Web mining Categories

Web mining can be defined as the discovery and analysis of useful information from the web document. It is categorized into three types that are Web content mining, Web structure mining, and Web usage mining. These three categories are described as follows and Fig. 2. Indicate the categories of web mining. Web content mining is the process of extracting useful information from the contents of web documents. It includes extraction of structured data from web pages, identify, match and integration of semantically similar data, opinion extraction from online sources, and concept hierarchy, ontology, or knowledge integration. Web content mining is the analogue of data mining techniques for relational databases since we can expect to find similar types of knowledge from unstructured data residing in web documents [5]. The content data corresponds to the collection of facts a web page was

resigned to convey to the users. It may consist of text, images, audio, video or structured records. Web content mining also distinguishes personal home page with other web pages.

Web usage mining analyses the transaction data, which is logged when users interact with the web. Web usage mining is sometimes referred to as 'log mining', because it involves mining the web server logs. Web server logs, which is maintaining an account of each user browsing activity. Web servers automatically, generate large data stored in server referred as logs containing information about the user profile, access pattern for pages, and so on [6]. The world's largest portal like Yahoo, MSN, and so on, needs a lot of insights from the behavior of their users' web visits. Web usage mining collects the data from web log records to discover users' access patterns of web pages. This can provide information that can be used for efficient and effective web site management and user behavior.
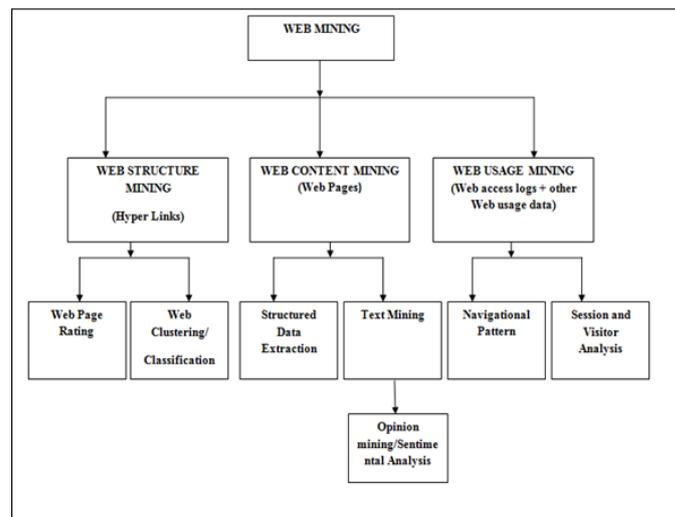

Fig. 2. Web Mining Categories

Web structure mining focuses on analysis of the link structure of the web and one of its purposes is to identify more preferable documents. The structure of a typical web graph consists of web pages as nodes and hyperlinks as edges connecting between two related pages [7]. Web page can also be organized in a tree-structures format, based on the various HTML and XML tags within the page. Technically, web content mining mainly focuses on the structure of the inner - document, while web structure mining tries to discover the link structure of the hyperlinks at the inter-document level. Based on the topology of the hyperlinks, web structure mining will categorize the web pages and generate the information, such as the similarity and relationship between different web sites [8]. The goal of web structure mining is to generate structural summary about the web site and web page.

II. RELATED WORK

Carmona et al. [11] determined perception of improving the E-commerce website design by web usage mining. Web usage mining defined as extracted the useful information from the user's history. This subsists of an E - commerce website called as orolivesur.com and also it was performed in a successful way by three phases. The phases were noted as the compilation and pre-processing of data, data mining, analysis and validation. A set of descriptive techniques is helping the webmaster team to improve the design of the website.

Divya Rajan et al. [12] discovered the technique that draw an elegant to improve the design of an e-commerce website design using web usage mining and k-means clustering. Web

usage mining is a technique that motivates website design to consumers in significant ways. It is used to provide the categorization of user's information based on transactional data. This clustering technique could improve the effectiveness of the websites by adapting the information structure of the sites to the user's behavior.

Suleiman et al. [13] analyzed customer loyalty in e-banking by structural equation modelling approach. It had become a benchmark for measuring success in a competitive economic environment where speed in innovation is the survival kit. Loyalty was the endogenous latent variable measured by satisfaction, reliability, responsiveness and empathy. Reliability was a significant positive antecedent of customer satisfaction via loyalty, then responsiveness to loyalty, satisfaction to loyalty showed significant relationships. Empathy, the model fit shows a mixed state, and thus it cannot be a representative for generalizations since studies are divided on positive and negative lines. The data used is sourced from one region; the model could not be generalized to the population.

Eric Hsueh-Chan Lu et al. [14] evaluated a framework for personal mobile commerce pattern mining and predicting of mobile users' movements and purchase transactions under the context of mobile commerce. The mobile commerce explorer used to mine the pattern and predict the user by similarity inference model, personal mobile commerce pattern mine, and mobile commerce behavior predictor. Mobile users move between the stores, the mobile information which includes user identification, stores and item purchased are stored in the mobile transaction database. This framework is to support the prediction of next movement and transaction.

Krithika et al. [15] concluded the prediction of m-commerce user behavior that used a weighted periodical pattern mining. This pattern mining used five components to measure the user's behavior that are a similarity inference model, mobile commerce behavior predictor, weighted mobile commerce behavior predictor, weighted mobile commerce behavior periodical predictor and performance evaluation. These components used to calculate the user's behavior in efficient manner this method helps to improve the business in the mobile commerce application.

Yuewen Liu et al. [16] explained the moderating effect of value uncertainty for online auction. In this method used a technique Meta-Analysis that effect of design in online auctions. It had mainly focused on three main findings in the moderating method that are the public reserve price has a positive effect on the auction price, secret effect reserve option has a positive and the buy-out option has a positive effect when auction price items are of low value. It identified value uncertainty as the key moderator of the relationship between auction design option and auction outcomes.

Prathyusha et al. [17] illustrated the online auction system that how frauds were detected by the machine-learned models. This model used to detect and reduce the frauds in online auction systems. More pro-active models used to detect the frauds and reduced the illegal ways, but that machine learned model more effective and found the frauds in quickly. This online framework was easy to extend in any other applications.

Xenia Papadomichelaki et al. [18] explained e-GovQual service for assessing e-government. This service through the internet was development of sites that better serve the citizen's needs. It is an e-Gov service quality model, which is conceptualized and then a multiple-item scale for measuring e-Gov service quality of governmental sites where citizens

seek either information or service is developed, refined, validated, confirmed and tested. e-government websites in many cases was fashion and could have an important impact on the citizens' view of the government.

III. RECENT TRENDS IN WEB USAGE MINING APPLICATIONS

With the rapid growth of World Wide Web (WWW), Web mining becomes more promising and popular topic in web research [9]. Web mining plays an important role in E-commerce website and E-services to understand, how their websites and services are used and to provide better services for their customers and users. It is used to mine the information over the web [10]. This paper discusses about few web applications and the methods used in the applications with web mining categories.

A. Websites Designing

A website is a collection of web pages that is document accessible via the World Wide Web on the internet. E-commerce website design revolves around plenty of things for business and consumers to communicate and conduct business [11]. It reduces delivery time and labour cost, thus it has been possible to save the time of both vendor and the consumer. In e-commerce sites the crucial design parameters are efficient navigation, search, along with speed and technically simple and basics [12].

B. Customer Loyalty for e-banking

In the web mining applications are used to improve the customer loyalty in web application.The E-banking includes familiar and relatively mature E-based products in developing markets, such as telephone banking, credit cards, ATMs, Internet banking and direct deposit. It also includes electronic bill payments and products, mostly in the developing stage, including stored-value cards and internet based stored value products. Electronic banking also known as electronic funds transfer (EFT), is simply the use of electronic means to transfer funds directly from one account to another, rather than by check or cash [13].

C. M-Commerce

M-commerce is a stand for mobile commerce. It is used to buying goods and their services over the internet. This communication is done through a wireless technology those devices such as handheld cell phones, tabs, and so on. The following are some of the examples mobile ticketing, information services, mobile banking and so on. [14]. Mobile ticketing is the process of where customers can order, pay for obtaining and validate tickets from any location and at any time using a mobile commerce application in mobile phones [15].

D. Online Auction

Electronic auction conducted through online. The marketmechanism by which buyers make bids and sellers place offers, characterized by the competitiveness and dynamic nature by which the final price is reached [16]. The internet provides an infrastructure for executing auctions at lower cost, and with many more involved sellers and buyers. Individual customers and corporations alike can participate in this rapidly growing form of e-commerce [17].

45

E. Assessing e-government service quality

E-Governance refers to the processes and structures that encompass all forms of electronic interaction between government and the citizens [18]. Through this application, it is used to share the useful communication to the citizens to government and government to citizens and also it provides access to information to empower citizens [19].
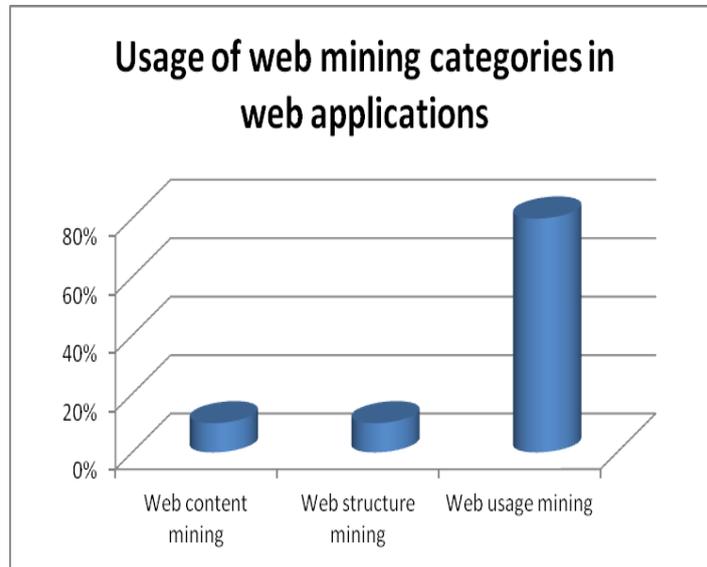

Fig. 3. Usage of web mining categories in web applications

TABLE I. RECENT TRENDS AND CHALLENGES IN WEB USAGE MINING APPLICATIONS AND ITS METHODS

S. No	Application	Techniques	Merits	Demerits
1	Improvement in the design of websites	Clustering, Association Rule	1. Usability 2. User Satisfaction 3. Tools are used to handle the data sets for real-time data.	1. Improved designs only for webmasters team. 2. Webmaster team can utilize the small number of data set with analytical tools.
2	Improving the elegant of Website design	K-Means Clustering	1. Effectiveness 2. Interesting patterns in large datasets.	1. It can extract only text files. 2. Information mining tools through only mining of web logs.
3	Customer loyalty for e-banking	SEM Approach	1. Satisfaction 2. Reliability 3. Responsiveness 4. Empathy	1. Measure the large amount of data set. 2. Generalizability is not possible to the population.
4	A Framework for personal m-commerce	SIM, PCMP-Mining, MCBP	1. Measuring among store and items 2. Efficient discovery of mobile user's 3. Predict the mobile user's behavior.	1. Low frequency in the similarity inference model 2. Low precision in predicting objects.
5	Predicting the user's behavior	SIM, MCBP, WMCBP, WMCBPP	1. More Efficient Service 2. More Accurate 3. Superior Performance	1. Lattice concept is not constructed for item proximity matrix and embeds it into kernel function. 2. There is no performing personalized service in the user space concept.
6	The uncertainty in online auction design	Meta-analysis method	1. Positive Effect. 2. The effect is stronger when the value uncertainty of auction items is higher.	1. No more moderators, 2. It does not estimate the most profitable level.
7	Detecting Frauds in Online Auction System	Machine-Learning	1. Reduces frauds and illegal activity 2. More Efficiently 3. More quickly than human tuned rule-based system.	1. Selection bias is not present in the online model training process and to deploy the online models for real production systems and also other applications.
8	Assessing e-government service quality	Multiple item scale	1. It measures the service quality distributed in government websites by using multiple item scale in efficiently with the four dimensions that are Reliability, Efficiency, Citizen Support, and Trust.	1. The service quality does not related to the electronic service stipulates to the populaces.

F. Web Advertisement

Web advertisement also referred to as an online advertisement. The use of popular websites can be an effective way of introducing new a product to the customer. Web advertisement referring to the internet and email based aspects of a marketing campaign, such a banner ad, e-mail marketing, search engine optimization, pay-per-click, and other tools [20].

G. Transaction Analysis

New environment brings new changes in the current economic model as it changes the relationship between operators and customers from the traditional physical store to electronic transactions on the internet. Analysis of e-commerce uses clickstream data to determine the marketing effectiveness of the site by quantifying user behavior while actually visiting the site visitor browsing the site recording the translation in a sales transaction [21].

H. Customer Relationship Management

Customer Relationship Management is becoming standard terminology, replacing misleading narrow term in relationship marketing [22]. It focuses creating value for the customer and company over the long term and the relationships are built with the customers, which provide value for services [23].

The above applications are used in day-to-day life by the users. The web mining categories are differ from each and every web applications. Most of the web applications are developed in web usage mining category and this is graphically discribed in Fig. 3. The discussed web applications are explained with its methods and categories in Table. 1.

IV. CONCLUSION

This paper discussed about the recent trends of web mining categories and chanllenges. The web applications are playing an important role over the internet in a different environment for all types of users. Most of the applications are used in the field of commerce. Now-a-days, the government is also using the web applications frequently like job portal, e-payments, certificate issuing, and identity cards. Today, people are doing most of their work in one place where they are sitting, and also the web applications reduce the time of folks in all ways.

REFERENCES

[1] Ashish Gupta, Anil Khandekar, "*Study Of Web Mining - A Survey*", ISSN: 2278 – 7798 International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 12, 2013.

[2] D. Jayalatchumy, P. Thambidurai, "*Web Mining Research Issues and Future Directions – A Survey*", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 14, Issue 3, PP 20-2, 2013.

[3] Brij Kishore, Dr. Raj Kumar, Prabhat Raghav, "*Research and Design of Web Data Mining*", Progress In Science in Engineering Research Journal, ISSN 2347-6680 (E), Vol.02, Issue: 02/06 March- April, Pages 311-315, 2014.

[4] Kavita Sharma, Gulshan Shrivastava, Vikas Kumar, "*Web Mining: Today and Tomorrow*", International Conference on Electronics Computer Technology, 978-1-4244-8679-3, VI 399-403, 2011.

[5] Niki R. Kapadia, Kinjal Patel, "*Web Content Mining Techniques – A Comprehensive Survey*", IJREAS, Volume 2, Issue 2, ISSN: 2249-3905, 2012.

[6] Namdev Anwat, Varsha Patil, "*Survey Paper on Web Usage Mining for Web Personalization*", International Journal of Innovative Research and Development, Vol 3 Issue 7, ISSN 2278 – 0211, 2014.

[7] Ahmad Tasnim Siddiqui, Sultan Aljahdali, "*Web Mining Techniques in E-Commerce Applications*", International Journal of Computer Applications", (0975 – 8887) Volume 69– No.8, 2013.

[8] D.Sridevi, Dr.A.Pandurangan, Dr.S.Gunasekaran, "*Survey on Latest Trends in Web Mining*", International Journal of Research in Advent Technology, Vol.2, No.3, E-ISSN: 2321-9637, 2014.

[9] S. Yadav, K. Ahmed, and J. Shekar, "*Analysis of Webmining Application and Beneficial areas*", IIUM Engineering Journal, Vol.12, No.2, 2011.

[10] Asha Joy, Remya, "*Techniques for Web Mining of Various Forms of Existence of Data on Web: A Review*", Internatioal Journals of Advance Research in computer science ans management studies, Volume3, Issue1, ISSN:2321-7782, 2015.

[11] C.J. Carmona, S. Ramírez-Gallego, F. Torres, E. Bernal, M.J. del Jesus, S. García, "*Web usage mining to improve the design of an e-commerce website: OrOliveSur.com*", Elsevier, Expert Systems with Applications 39 , 11243–11249, 2012.

[12] S. Divya Rajan, Mr. Neelabh Sao, "*An Elegant Draw Near to Improve the Design of an E-commerce Website Using Web Usage Mining and K-Means Clustering*", IJCSMC, Vol. 3, Issue. 6, pg.256 – 265, ISSN 2320–088X, 2014.

[13] Suleiman G. P, Nik Kamariah Nik Mat, Adesiyan O. I, Mohammed A. S, Jamal ALekam, "*Customer Loyalty in e-Banking: A Structural Equation Modelling (SEM) Approach*", American Journal of Economics, Special Issue: 55-59 DOI: 10.5923/j. economics. 20120001.13, 2012.

[14] Eric Hsueh-Chan Lu, Wang-Chien Lee, and Vincent S. Tseng, "*A Framework for Personal Mobile Commerce Pattern Mining and Prediction*", IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 5, 2012.

[15] S. Krithika, M. Moorthi, "*Prediction of M-Commerce User Behavior by a Weighted Periodical Pattern Mining*", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 6, 2013.

[16] Yuewen Liu, Kwok Kee Wei, Huaping Chen, "*A meta-analysis on the effects of online auction design options: The moderating effect of value uncertainty*", Elsevier, Electronic Commerce Research and Applications 9, 507–521, 2010.

[17] K. Prathyusha, T. Anuradha, R. Sai Nikitha, K. Meghana, "*Detecting Frauds In Online Auction System*", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 4, 2013.

[18] Xenia Papadomichelaki, Gregoris Mentzas, "*e-GovQual: A multiple-item scale for assessing e-government service quality*", Elsevier, Government Information Quarterly 29, 98–109, 2012.

[19] France Be´langer, Lemuria Carter, "*Trust and risk in e-government adoption*", Elsevier, Journal of Strategic Information Systems 17, 165–176, 2008.

[20] Yasser Lamlili El Mazoui Nadori, Mohamed Erramdani, Mimoun Moussaoui, "*Semantic Web-Marketing 3.0: Advertisement transformation by modelling*", IEEE International Conference on Multimedia Computing and Systems (ICMCS), 978-1-4799-3824-7, 2014.

[21] Andrew J. Flanagin, Miriam J. Metzger, Rebekah Pure, Alex Markov, Ethan Hartsell, "*Mitigating risk in eCommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention*", Springer Science+Business Media, Electron Commer Res 14:1–23, DOI 10.1007/s10660-014-9139-2, 2014.

[22] E.W.T. Ngai, Li Xiu, D.C.K. Chau, "*Application of data mining techniques in customer relationship management:A literature review and classification*", Elsevier, Expert Systems with Applications 36, 2592–2602, 2009.

[23] Moutusy Maity , Mayukh Dass , "*Consumer decision-making across modern and traditional channels:E-commerce, m-commerce, in-store*", Elsevier, Decision Support Systems 61, 34–46, 2014.