

A REVIEW PAPER ON IDENTIFYING STUDENTS INTEREST IN E-LEARNING USING WEB USAGE MINING

Sunil¹ and Prof. M. N. Doja²

Abstract--In the present era, World Wide Web has become a popular and unavoidable medium to extract useful and relevant information from the huge and scattered information particularly for education and e-learning. In every educational organization e-learning is becoming an integral part of teaching-learning process. Web mining is the use of data mining techniques to automatically discover and extract relevant and useful information from World Wide Web documents and services. The Web usage mining includes the data from the web server logs, proxy server logs, browser logs, user profiles, registration data, user sessions or transactions, cookies, user profiles, registration data and any other data as the results of interactions. This paper gives the overview of various methods of e-learning and aims to experiment with pattern classification for students' behaviour and their interest in e-learning. It then identifies the problems within e-learning that web usage mining can solve as well as present the existing approaches.

Keywords— Data mining, web usage mining, server log, e-learning.

I.INTRODUCTION

E-learning is a structure that consists of all form of electronically supported learning and teaching at anytime and anywhere. The communication and information plays an important role in implementing the learning process.

E-learning is a critical support mechanism for educational institutions to grow the performance of their students, teachers, as well as useful for organizations to enhance the performance of their employees [2]. E-learning aims to deliver individualized, comprehensive, dynamic learning content in real time, aiding the development of communities of knowledge [3]. E-learning offers new possibilities in learning. Thus a student can get immediate feedback on solution to problems, learning paths can be individualised, etc. E-learning is an alternative concept to the traditional tutoring system with the application of more and better computer techniques in computer.

The alarming experimental rate at which the World Wide Web is growing in both the sheer volume of traffic and the complexity of website, it has become very important to analyze this web traffic and the usage of the web sites by the users. This growth of the World Wide Web has lead to develop different client side and server side tools that mine the information resources to extract knowledge. Nowadays, World Wide Web has become a popular medium to extract useful and relevant information from the huge amount of information particularly for education and e-learning. In every educational organization, e-learning is becoming an integral part of the learning. An opulence

¹ *Department of Computer Engineering Faculty of Engineering & Technology Jamia Millia Islamia New Delhi, India*

² *Department of Computer Engineering Faculty of Engineering & Technology Jamia Millia Islamia New Delhi, India*

class of reusable educational resources [1] such as electronic tutorials, textbook, digital libraries and lot more other huge amount of repositories of educational materials are being offered by the World Wide Web.

The paper is organised as follows. Section II elaborates on the various e-learning tools and techniques while section III talk about web usage mining section IV presents the various sources of web mining related work. Section V talk about web usage mining process, section VI is about related work and section VII concludes the article with comments on future trends.

II. OVERVIEW: E-LEARNING TOOLS AND TECHNIQUES

E-Learning is becoming an important tool to support the learning system to achieve its goals. E-Learning became hot topic in the 1990's after the spread of the Internet. Although it has a relative short history, it is becoming an important part of the learning [4]. The majority of the universities adopted some kinds of e-learning within its learning system. E-learning may either be synchronous or asynchronous[19].

Several tools, both general purpose such as database, blogs, Wikis, etc. as well as specific such as Moodle, Blackboard, etc. are used in the context of e-learning [5].

V. Sklyarov et al. [6] proposed advanced e-learning tools and remote laboratories for state-of-art engineering education and discussed about importance of reconfigurable systems in education and proposed effective learning methods and tools, that includes laboratory templates, animated tutorials, education oriented examples, remotely reconfigurable prototyping system, and virtual software hardware co-simulation environment.

S. Maria [7] proposed e-learning system development for engineering education and discussed about the availability and quality of e-learning, individualization and integrity of learning, need of advanced technologies. Also discussed about various trends to use cloud computing technology in which computer resources and facilities are available to the user as a web service in the sphere of education. How could this impact engineering education, so as to supply better result of e-learning and enforce its accessibility, diversity and powerfulness.

J. Chen et al. [8] designed and implemented a new e-learning system, which incorporated three innovative internet technologies as easy knowledge search and classification, instant chat system and Ideone online compiler system. This new e-learning system is based on a well known e-learning system named Chamilo. Chamilo is an open-source e-learning and content management system, aimed at improving access to education and knowledge globally.

R. Cação et al. [9] proposed a model of quality in e-learning that composed of three factors. This model implies the perceptions of quality in e-learning can be explained, with comfortable goodness-of-fit, by three factors: the training process, the training attitudes, and the training utilities. The Structural Equation Modelling (SEM) is used there to test and confirm the structure of a model of critical factors of quality in e-learning.

III. WEB USAGE MINING

Web mining is the application of data mining techniques to extract knowledge from web data. Data mining is the process of extracting knowledge and useful hidden information from a huge amount of incomplete and noisy data. Web data mining is the process that applies data mining technique to these web data bases. Web data mining broadly divided into three distinct categories, according to the kinds of data to be mined, that are web content mining, Web structure mining and Web usage mining. Figure 1 shows the taxonomy.

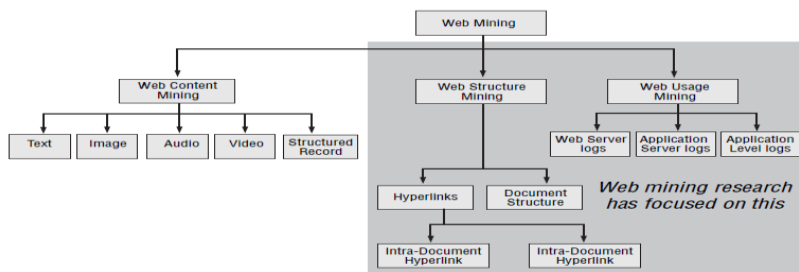


Figure 1: Web Mining Taxonomy

- **Web Content Mining**

Web content mining is the process of extracting useful information from the contents of web pages. Content data is the collection of facts that a web page is designed to contain. Content data consist of text, image, audio, video, hyperlink or structured records such as lists or tables.

- **Web Structure Mining**

Web structure mining is the process of obtaining structure information from the web. It generates structure summary of the web page which is known as web graph. The structure of web graph consists of web pages as nodes and hyperlinks as edges connecting related pages. It is further divided into two kinds based on the kind of structure information used. i.e. Hyperlinks and Document Structure.

- **Web Usage Mining**

Web Usage Mining is the process of extracting knowledge or useful information from web usage data, in order to understand and better serve the needs of web based allocations [20].

IV.SOURCE OF DATA FOR WEB USAGE MINING

The data is the basics of a knowledge discovery process. As it is shown in figure 2 there are several possible data sources for the Web usage mining process. Each type has its own advantages and a little different focus. There are three data sources, they are:

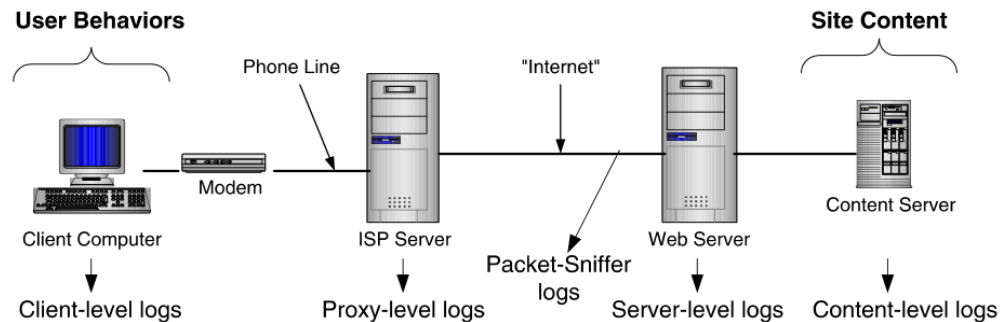


Figure 2: Data Sources of different web logs.

A. Server Level Collection

Server level data collections are often used because of their easy availability – by default every Web server saves logs. For Web Usage Mining process, Web server logs are the best known data source. The log usually contains information about the remote host name or IP address, the user name, requested URI, server status, time and date, and transferred bytes. The Extended Log format adds other two entries: referrer and user agent identification which are very useful for session identification.

Similar work as Web server logs is done by packet sniffers which monitor network traffic by watching server incoming and outgoing TCP/IP packets. Their advantages over Web server logs are that they can save POST parameters of CGI scripts and HTTP headers of requests. Another server level data sources are logs from content (or application) servers which serve contents for dynamic Web sites. Their advantage is that they can provide information about all incoming parameters, internal state variables, and outgoing data.

B. Client Level Collection

Client level data collections are focused on tracking single user sessions across single or multiple Web sites. Their main disadvantage is that they are based on user cooperation. There are two ways – remote agents (JavaScript or Java applets) and modified browsers. The first way is aimed at single user sessions across a single server. It solves the problem with caching and almost with session identifying. The best results for single users or multiple sites are given by special Web browsers that track every user movement. The browser records how much time the user

spends on a Web page, if he pushes the back or reload button, and many other valuable variables. But it is hard to persuade users to use such a special browser.

C. Proxy Level Collection

The last data source is proxy server logs. It acts as an intermediate level of caching between client browsers and Web servers. Proxy server logs save multiuser or multisite communication and are suitable for characterizing the browse behaviour of a group of users sharing the same proxy server.

V. WEB USAGE MINING PROCESS

Web usage mining is the application of data mining techniques to discover interesting usage patterns from web usage data, in order to understand and better serve the needs of web-based applications [10]. Usage data captures the identity or origin of web users along with their browsing behaviour at a web site. Figure 3 shows web usage mining process. The Web usage mining is parsed into these distinctive phases/stages:

A. Data Collection and preprocessing

Within this stage, usage data from various sources are collected from web servers, clients connected to a server, or from middle sources such as proxy servers and client side, browser based application. The data preprocessing of Web usage mining is usually complex. Purpose of data preprocessing is to offer reliable, structural and integrated data source to pattern discovery. It consists of four processes: Data cleaning, User identification, Session identification, Path completion [11].

B. Pattern Discovery

Pattern discovery is the key process of the Web mining, which covers the algorithms and techniques from several research areas, such as data mining, machine learning, statistics and pattern recognition. The techniques such as statistical analysis, association rules, clustering, classification, sequential pattern and dependency modelling are used to discover rules and patterns. The knowledge that can be discovered is represented in the form of rules, tables, charts, graphs, and other visual presentation forms for characterizing, comparing, predicting, or classifying data from the Web access log.

C. Pattern Analysis

The final stage of the whole Web usage mining is pattern analysis. The goal of this process is to extract the interesting rules or patterns from the output of the pattern discovery process by eliminating the irrelative rules or patterns. There are two most common approaches for the patter analysis. One is to use the knowledge query mechanism such as SQL, while another is to construct multi-dimensional data cube before perform OLAP operations [12].

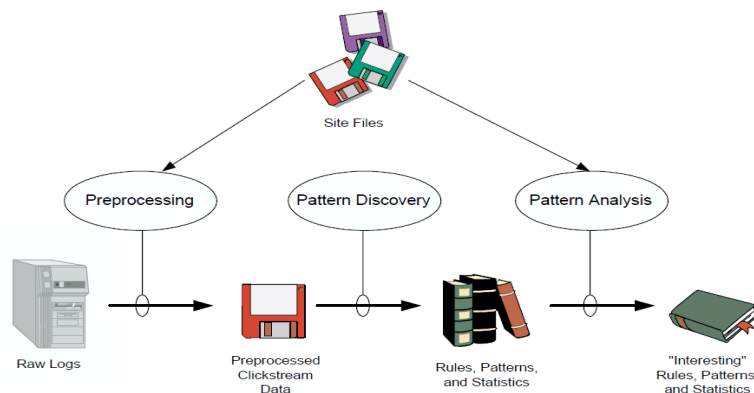


Figure 2: Web Usage Mining Process

VI. RELATED WORK

As e-learning similar to e-commerce is a large and growing business the Web usage mining is considered as a suited candidate to fit into the domain of e-learning [5]. The primary goal of applying web usage mining techniques in e-learning is to deal with the exponentially growing and unstructured information the e-learning systems are generating. I was conceived that the web usage mining techniques would enable data analysis to transform the sheer volume of data to relevant contextual knowledge.

Xinjin Li et al. [13] proposed a model of e-learning platform with learning resources recommendation based on web usage mining and used analog and SQL Server 2005 to mine the log files on the server. It combines the web usage mining technology and e-learning platforms to build a network learning platform model with function of recommending learning resources.

Péter Tóth [14] proposed his work for adaptive online learning environment and web usage mining. There by the application of web mining method it was tried to find an answer to the question whether any conclusions from the patterns of online learning activities were to be drawn as to preferred learning characteristics, methods and strategies and also whether the major variables of online learning behaviour were possible to define. In another proposed work by Péter Tóth [15] some useful information is collected by analysing the students' behaviour as to develop a model that is based on the systematic instructional design model (ADDIE) and Kirkpatrick's general evaluation model. That results to make it possible to analyse students' behaviour informally and quality assurance.

W. Chandrama et al. [16] proposed a survey on data preprocessing method of web usage mining. There discussed about the different algorithms and methods which are available for data preprocessing for analyzing user navigational pattern the methods used are like Correlation Regulation Discovery, Sequence Pattern Recognition and Cluster pattern. The algorithm, K-mean is used for Clustering and Apriori is used for Correlation Regulation.

Xue Sun et al. [17] proposed an e-learning model based on web usage mining techniques and implies the model through new algorithms. This model benefitted in the increase of the intelligent study and promotes the interests of learners.

A. Yadav et al. [18] proposed the effectiveness of the WUM techniques that can make e-learning environment more efficient. It was suggested by using a model with the implementation based on web usage mining technique. In a result building effective learning system and promote the interest of learners.

VII.CONCLUSION

Web usage mining is suitable and active concerning field of research that effectively used to enhance e-learning platform that leads to develop intelligent and personalised environment or learners. This paper has provided a more current evaluation and updates of e-learning and web usage mining techniques. E-learning is becoming an important tool to support the learning system and offers new possibilities in learning. Web usage mining techniques within e-learning systems is a rapidly growing phenomenon. Based on the analysis of the concept and process of the web usage mining, this paper combines the web usage mining technology and e-learning platforms.

Among them predictive modelling has particularly interesting applications on predicting students' performance and the usefulness of e-learning systems. There are currently many open research challenges that exist in this domain, particularly centred around data privacy and ethics.

REFERENCES

- [1] A. Lalingkar, S. Ramani, "A Students Assistant for Open E-Learning", International Workshop on Technology for Education (T4E), Aug 4-6, 2009, Bangalore, India, pp. 62-67.
- [2] A. Martin, C. León, "An intelligent e-learning scenario for knowledge retrieval". IEEE.
- [3] Liu MinLi, Yu Chao, Mu Dan, "The Method of Dynamic Services in E-learning System", International Forum on Information Technology and Application, 2010, pp. 259-262.
- [4] Dr. Jamil A. Itmazi, "E-Learning Systems and Tools" 1st editing, Publisher: Phillips Publishing, on April 2010, ISBN: 978-0-557-44966-8.
- [5] Yousuf A. ALMazroui, "A survey of data mining in the context of E-learning", International Journal of Information Technology & Computer Sciences (IJITCS), Jan-Feb, 2013, Volume 7: No:3.

A Review Paper On Identifying Students Interest In E-Learning Using Web Usage Mining

- [6] V. Sklyarov, I. Skliarova, "E-learning Tools and Remote Reconfigurable Systems for Engineering Education", International Conference on Networking and Information Technology, 2010, pp. 11-15.
- [7] S. Mariya, "Advanced Technologies of E-learning in Engineering Education", 14th International Conference on Interactive Collaborative Learning (ICL2011), 21-23 September, 2011, Slovakia, pp. 565-568.
- [8] J. Chen, Q. Li, C.Y.K Lin, H. Chang, C. Wang, "Application of Innovative Technologies on the E-learning System", The 6th International Conference on Computer Science & Education (ICCSE 2011), August 3-5, 2011, SuperStar Virgo, Singapore, pp. 1033-1036.
- [9] R. Cação, C.L de Freitas, "Modeling Critical Factors of Quality in e-Learning", 12th IEEE International Conference on Advanced Learning Technologies, 2012 pp. 422-426.
- [10] R. Cooley, "Web Usage Mining: Discovery and Application of Interesting Patterns from web data", PhD thesis, Department of Computer Science, University of Minnesota, May 2000.
- [11] Arvind Kumar Dangi and Sunita Sangwan, "A new Approach for User Identification in Web Usage Mining Preprocessing", IOSR-JCE, May 2013.
- [12] O. Zaiane, M. Xin, J. Han., "Discovering Web Access Patterns and Trends by applying OLAP and Data Mining Technology on Web Logs". In *Advances in Digital Libraries*, pages 19-29, Santa Barbara, CA, 1998.
- [13] Xinjin Li, Sujing Zhang, "Application of Web Usage Mining in e-learning Platform", International Conference on E-Business and E-Government, 2010, pp. 1391-1394.
- [14] Péter Tóth, "Adaptive Online Learning Environment and Web Usage Mining", 8th IEEE International Symposium on Applied Computational Intelligence and Informatics, Timisoara, Romania, May 23-25, 2013, pp. 61-66.
- [15] Péter Tóth, "E-Learning and Web Mining", IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics, September 20-22, 2012, Subotica, Serbia, pp. 439-444.
- [16] Wasvand Chandrama, Prof. P. R. Devale and Prof. Ravindra Murumkar, "Survey on Data Preprocessing Method of Web Usage Mining", International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 5 (3), 2014, pp. 3521-3524.
- [17] Xue Sun, Wei Zhao, "Design and Implementation of an E-learning Model Based on WUM Techniques", International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government, 2009, pp. 248-251.
- [18] Anuradha Yadav, Satbir Jain, "Analyses of Web Usage Mining Techniques to enhance to Capabilities of E-learning Environment", IEEE, pp. 223-225.
- [19] M. M. Shahabadi, M. Uplane, " Synchronous and asynchronous e-learning styles and academic performance of e-learners", *Procedia-Social and Behavioral Sciences* 176(2015), pp. 129-138.
- [20] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", SIGKDD Explorations, ACM SIGKDD, Jan 2000, pp 12-23.