



Web Usage Mining Techniques to Improve the Capabilities of E-learning Websites and Blogs

Sunil

Section of Computer Engineering
Department of University Polytechnic
Faculty of Engineering & Technology
Jamia Millia Islamia, New Delhi, India

Prof. M. N. Doja

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

Abstract: The current scenario of advanced technologies and rapid development of internet and its efficiency learners prefer online learning. A new way of education system is created with the rapid growth of internet and learners realize their E-learning activities with less effort, time and money. Now days internet is most important media for the collection, distribution and sharing of information. E-learning is the area where web-based technologies have been rapidly and effectively adopted. By using various data mining techniques the transformation and interpretation of web data is done to find out the various hidden patterns and information which results in enhancing E-learning environment. For the E-learning environment web usage mining is used to mine the server logs to find the learners' usage pattern so that the learners can be provided with an efficient platform and more personalized services. The paper introduces the use of web usage mining techniques that makes the E-learning environment more competent and effective. With the use of web usage mining techniques the websites and blogs must be designed such a way that it satisfies the needs of the learners and provide more effective learning environment. The maintaining and restructuring of websites and blogs which in turn helps developers to increase the visits of current learners and promote interest of learners also attracts new ones.

Keywords: E-learning, web usage mining, personalized, learners

1. INTRODUCTION

Fast developments in the communication and information technologies have fuelled the extensive use of internet, which has changed the learning behaviours of the learners. The internet has become an information hub that is accessible in a pervasive manner. Now days it is an admirable platform for delivering and acquiring information for learners. E-learning systems engage users in a process of learning by providing then necessary useful information on-time and in effective way.

The term E-learning is generally referred as computer-enhanced learning [1], and the first general purpose E-learning system was Programmed Logic for Automatic Teaching Operations (PLATO) which was developed by the Computer-based Education Research Laboratory (CERL) of University of Illinois at Urbana Champaign in early 1960's. The well-known quality of E-learning system is that the students become self-learner and explorers of useful information [2]. In the learning process, the learner is the main body but not inactive receiver. E-learning system provides all kinds of learning manners to the students to comprehend interactive, real-time and accommodating learning at different places. To grow up knowledge of students in every field of study most of the educational institutes are using E-learning based systems and provide interactive systems to the students which makes communication faster [3]. There are many advantages of adopting the E-learning systems as paper less environment, learners can access the information from any part of the world, and advanced study material is provided and can be accessed from homes.

Billions of users visiting E-learning websites and blogs for learning and other purposes they leave an enormous amount of information stored in log files [4]. The hidden knowledge in the data is analysed and can be used to improve the E-learning websites and blogs. For

improvement of E-learning websites and blogs the web usage mining techniques are applied on web based learning systems. This technique can analyse, classify and cleanup the mining result to offer suitable resources to every learner.

2. WEB MINING

Web mining is one of the important branches in data mining also we can say web mining is the application of data mining techniques to extract knowledge from web data. Web data mining is the process that applies data mining technique to web data bases. Web mining is extraction of useful information or pattern from the web data. It can be classified into following three distinct categories according to the kind of data to be mined, that are web content mining, Web structure mining and Web usage mining as shown in figure1 [8]:

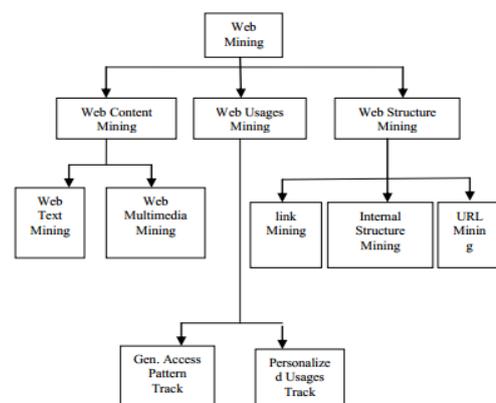


Figure 1: Web Mining Taxonomy

A. Web Content Mining

The process of extracting valuable information from the contents of web pages is web content mining. Content data is the compilation 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.

B. Web Structure Mining

The process of obtaining structure information from the web is Web structure mining. 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. Based on the kind of structure information used, it is further divided into two kinds. i.e. Hyperlinks and Document Structure.

C. 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 [10]. Web usage mining refers to the automatic detection and analysis of patterns in click stream and related data collected or generated as a result of learners interactions with web resources like websites and blogs. The major objective is to capture, model, and analyze the behavioural patterns and profiles of the learners how are interacting with the website and blogs. The revealed patterns are usually represented as collections of pages, objects, or resources that are frequently accessed by groups of learners with common desires or interests.

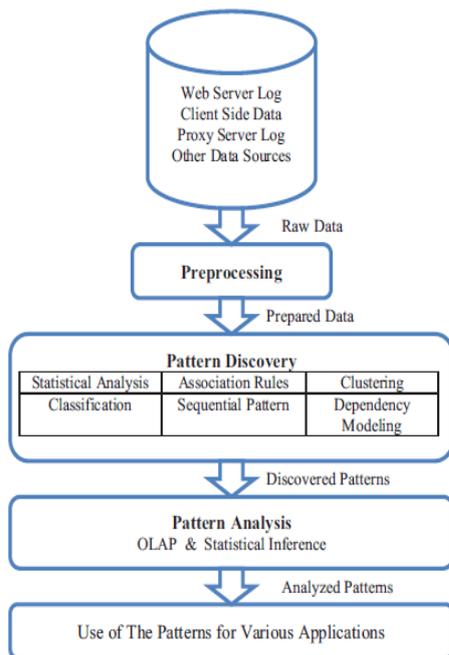


Fig. 2. Web Usage Mining Process

3. PROCESS OF WEB DATA MINING

The Web data mining process includes: Data Collection, Data Pre-processing, Data Storage, Data Mining and the finally Knowledge discovery is done [5].

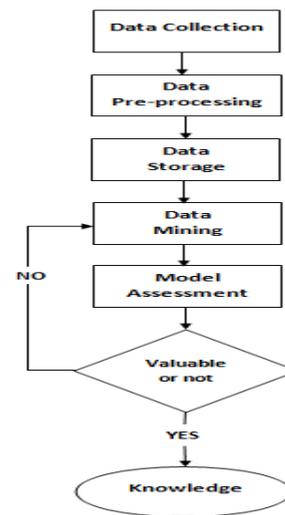


Fig. 3. Process of web data mining

A. Data collection

The first step is to collect the data on which we want to discover the hidden information. The data source of the e-commerce website includes customer personal information; customer personal information includes not only the customers' registered personal data but also includes its order information, customers' individual requirement and problems. Also data is collected about customers browse records and visited pages and behaviour information that help to analyse customer preference so as to forecast the future purchase behaviour of the customers. The second source is the server information that is generated on the server when client access the server, the data includes server log, error logs, agent logs files, cookie logs and transaction database.

B. Data pre-processing

As after the data is collected as log files the pre-processing is being done. The data pre-processing includes data cleansing, user identification and transaction identification.

a) Data cleansing

In data cleansing the unwanted data or we call errors are removed and the deletion of redundancy in data is done. The data cleansing done on multimedia files, java applet files, java script files, CSS files, pop-up ads and error access records.

b) User identification

In this every user in being identifies from the use log files. The most common technique used is based on the idea of heuristic rules as the different IP addresses represents different users, although id and the IP addresses are same then the agent log files can figure out whether the user is same or not by detecting the browser and operating system, if there is a condition that IP address, browser and operating system are same, then the user can be identified by page link structure.

c) Transaction identification

In this the users' behaviour is defined from their records. Here first each record is described then identified sessions of every user visit is organised in chronological order.

C. Data Storage

After data pre-processing is done the data now stored in the database in order to get ready to be extracted and used.

Here in web data mining transactional database is used as compared to relational database.

D. Data Mining

After pre-processing and data storage we choose different data mining technologies that are based on the different requirements of the users. There are many data mining technologies which are association rules, clustering, classification and sequential patterns.

a) Association rules

By applying the association rule generation, in the web domain the pages which are largely often referenced together can be put in one single server session. To discover unordered relationship between items found in a database of transactions, association rule mining technique can be used[7]. In terms of web usage mining the association rules refer to a set of pages that are being accessed together with a support value that exceeding some specified threshold[6]. With the presence or absence of the association rules the web designers can restructure their web sites efficiently.

b) Clustering

It is a technique to group together the data items or users with similar characteristics. Clustering of user information or data items can facilitate the expansion and carrying out of future marketing strategies [7]. The users who have similar navigation patterns, group of those users can be revealed by clustering.

c) Classification

It is the technique of mapping a data item into one of the several predefined classes. In the web domain if users belonging to a particular classes or category, by using this technique web master can create a profile of such users. This requires extraction and selection of features that best describes the properties of a given class or category. By using supervised inductive learning algorithms such as decision tree classification, k-nearest neighbour classifier, support vector machine, naïve Bayesian classification etc, the classification can be done [6].

d) Sequential Pattern

The sequential pattern techniques aim to find the inter-session pattern, as a set of the items follows the presence of another in a time-ordered set of sessions. For the web marketer it is very meaningful to predict the future trends.

E. Model Assessment

When the system has implemented the algorithm, we get a huge number as thousands of models or rules. Only a small part of them is valuable and interesting. So for that the explanation and assessment is essential. To understand those complex data through using computer as to create some visible images, the technology used is visualization. As some decision makers are not so expert in data mining, that's why these models are required to convert into the forms as they are easier to understand and used by the users to make decisions.

4. RELATED WORK

Web usage mining is one of the best areas of researches now days. Web usage mining technique being used widely to discover the learners' navigation patterns from web server logs [9]. A well designed website is always considered as a

successful website. The website must be designed such a way that it can convince the requirements of the users.

Lya Hulliyyatus Suadaa [11] has done a survey on web usage mining that conducted with the Systematic Literature Review method to identify relevant studies about data sources, applications, techniques and existing issues. A Systematic Literature Review is an approach that is to recognize, assess and construe about all the related studies about a particular research question, topic area or fact of concern.

S. Prakasam et al. [12] proposed an architecture model of E-learning process using the agent concept. The Agent Based Intelligent System (ABIS) is anticipated to improve the E-learning also simplifying and automating the process of creating the domain model for an intelligent E-learning system. The ABIS is a method which provides a direct customized directions and feedback to the learners without the intrusion of human beings. This Agent based recommendation system helps communities of learners in searching the web for their desired useful information.

P. Veeramuthu et al. [13] proposed method for enhancing the learning capabilities of the learners in educational institute. The study aims to analyse about the different factors that are affecting a learners learning behaviour and their performance using clustering for the data mining techniques. Here clustering analysis is used to part learners into groups according to their features. As clustering is the one of the basic technique frequently used in analyzing the data sets and clustering considered the most vital unsupervised learning technique. It is used to segment a large set of data into subsets. Each cluster is group of data objects that are comparable to other cluster placed within the same cluster but unlike to object in other cluster. Here the data gathered from college students is analysed using a data mining technique that is k-mean clustering to predict the students/learners result so as to enhance their learning capabilities.

Xue Sun et al. [14] proposed a learning platform that aims to build an individual learning platform for every learner. The web mining technique is used here in digging web logs. As it can rapidly find out users browsing patterns and provide intelligent services to the learners. The proposed individual learning platform is introducing web mining to pertain web-based learning systems. The system can explore, categorize and cleanup the mining result sets and provide a suitable schema for every learner. The web usage mining technique as one of the tangible technique of web mining that can extract the useful information about learners' accesses obtained from web server logs. It can be very helpful to build an intelligent and individual E-learning platform for learners by analyzing the extracted information.

S. Maria [15] proposed an e-learning system for the engineering education and discussed about the accessibility and excellence of e-learning platforms, individualization and reliability of learning platforms, need of advanced technologies. Also it has been discussed about diverse trends to use cloud computing technology in which computer resources and amenities are available to the learner as a web service in the sphere of education and learning so as to deliver better result of e-learning and inflict its accessibility, diversity and powerfulness.

J. Chen et al. [16] designed and implemented a new E-learning system. This new E-learning system is based on a

well known E-learning method named Chamilo, which integrated three modern internet technologies as easy knowledge search and classification, instant chat system and Ideone online compiler system. Chamilo is an open-source E-learning and content management system, aimed at recuperating access to education and knowledge worldwide.

R. Cação et al. [17] proposed a model of excellence in e-learning that composed of three factors. This model implies the perceptions of excellence 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 analysis and confirms the composition of a model of critical factors of excellence in E-learning.

5. THE PROPOSED SYSTEM

The proposed system is based on the method proposed by Rana Khudhair Abbas Ahmed [18]. In this work, the system named COMBI is designed to saves information about the learners learning requirements without asking them openly. Based on the learners' preferences, educational background and experience, this information helps the system to make personalization of the learning content and the information is stored in the learners' profile. The anticipated system combines three web mining techniques (web content mining, web structure mining, web usage mining). The architecture of the system is depicted in figure 4.

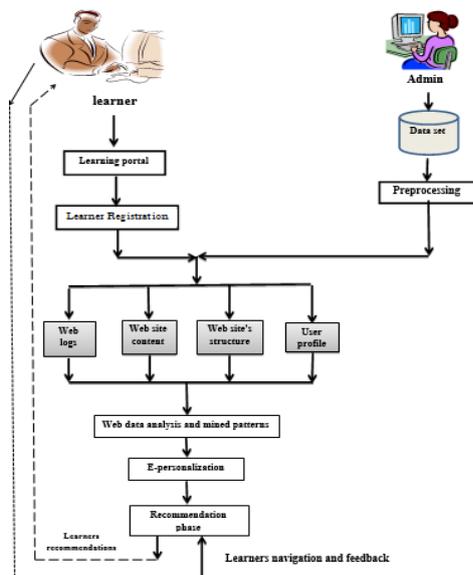


Fig. 4. The architecture of the proposed system

Following are the major modules of the system.

A. Learning Portal

If the learner is new, then first of all the learner have to make registration to the learning portal.

B. Learners Registration

In this Learners Registration process it saves the personal information about the learner in a profile. When the learner login to the system using its user name and password, then the learner can makes search for the desired information.

C. Admin

Admin is responsible for forming a standard dataset including learning objects of the learners.

D. Data Sources

The data can be obtained from various sources like:

- Usage Data: The log data collected involuntarily by server that represents the navigational behaviour of the learners. Each hit, represents a single entry in the log record. Each log entry consists of all information like time, date of request, IP address of the learner, the resource requested, mode of request, HTTP status, cookie etc. This data needs to be transformed and aggregated at different levels of abstraction.
- Content Data: It includes static HTML/XML pages, multimedia files and compilation of records from operation databases.
- Structure Data: Forms the structure data web designers vision of content organization, within the site e.g. HTML/XML pages can be represented as a tree structure.
- User Data: It includes profile data and the demographic information about registered learners.

E. Preprocessing

It is the first step in Web Mining. It is necessary to convert the data to and suitable form for solving a definite educational problem. This includes choosing what data is to collect, focusing on the questions to be answered and making sure the data support with the questions. This is useful to retrieve this appropriate data set from raw web log records, to which various data mining & statistical techniques could be applied.

F. Learner

A learner profile is generated from learner allied information; this includes learners' preferences, learners' demographics learning knowledge, and learning styles. This system generally deals with unstructured data related to the items.

G. Web Data Analysis and Mined Patterns

After data cleaning, filtering, pre-processing, integrating from multiple sources, the integrated data is transferred into a warehouse, suitable to be used as input to various data mining techniques with the learner's profile. Interesting patterns could then be mined. For learner identification, one needs to identify who accessed website and what pages were accessed.

H. E-Personalization

In the E-learning personalization process, there is much information available about not only learners' interface rather about actions, such as reading, writing, taking assignments and communication with peers. Analyzing the server logs and the history list can help to understand the learners' behaviour and the web structure, thus recuperating the design of the website.

I. Recommendation phase

Web mining methods were used to incorporate previous learners' feedback for making recommendations. The proposed system uses the learners' profile to distinguish their goals and envisage a recommendation list. The recommendation procedure is performed by mining learners' profile, comparison of extracted characteristics from hidden items with content similes in the user profile of learners, extracting learner preferences from the learners active session and computing significant links to recommend for the active learner.

J. Learners' navigation and feedback

In this the learner is provided with a list of recommendations and is required to prefer one of the

recommendations that best suit their requirement. The system then uses this feedback to present the learner with other similar objects. The iterations continue until the learner finds an object of interest or abandons the search. After discovering the learning behaviour pattern the system can build up a series of feedback and motivations to reach the learners satisfaction.

6. CONCLUSION

Web usage mining is suitable and active concerning field of research that efficiently used to enhance E-learning system that can be benefit to develop intelligent and personalized environment for the learners. E-learning is becoming a vital tool to sustain the learning system and offers new potential in learning. Web usage mining techniques within E-learning systems is a rapidly growing trend. Web mining is the series of task used for mining or extracting useful information from the web pages or websites. Based on the analysis of the concept and process of the web usage mining, this paper combines the web usage mining techniques and E-learning platform. Here the proposed system saves information about the learners' learning requirements without asking them unambiguously. This information helps the system to make personalization of the learning content based on the learner's preferences, educational background and experience. This system can increase the intelligent study and promote the interest of learners along with predicting learners' performance and the usefulness of E-learning systems. There are currently many open research challenges that exist in this domain, particularly centred on data privacy and ethics.

7. REFERENCES

- [1] Chien-Chung Chan, "A Framework for Assessing Usage of Web-Based E-learning Systems", IEEE 2007, 0-7695-2882-1/07.
- [2] Manasi Chakurkar, Prof. Deepa Adiga, "A Web Mining Approach for Personalized E-learning System", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 5, No. 3, 2014. pp. 51-56.
- [3] R. Karthiya Banu, Dr. R. Ramanan, "Analysis of E-learning in Data Mining- A Dreamed Vision for Empowering Rural Students in India", IEEE-International Conference on Recent Trends in Information Technology (ICRTIT), MIT, Anna University, Chennai, June 3-5, 2011, pp. 1265-1270.
- [4] Anuradha Yadav, Satbir Jain, "Analysis of Web Usage Mining Techniques to Enhance the Capabilities of E-learning Environment", IEEE 2011, pp. 223-225.
- [5] Yanduo Zhao, "The review of web mining in E-commerce", International Conference on Computational and Information Sciences, IEEE 2013, doi:10.1109/ICCIS.2013.158, pp.571-574.
- [6] 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.
- [7] R. Cooley, B. Mobasher, J. Srivastava, "Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, 1(1), 1999.
- [8] Sunil, Prof. M.N. Doja, "A Review Paper on Identifying Students Interest in E-learning using Web Usage Mining", International Journal of Latest Trends in Engineering and Technology (IJLTET) Volume 8, Issue 1, pp.520-525, e-ISSN:2278-621X, doi:http://dx.doi.org/10.21172/1.81.067.
- [9] Sunil, Prof. M.N. Doja, "Recommender System Based On Web Usage Mining For Restructuring Of E-Learning Websites And Blogs", International Journal of Research in Engineering & Advanced Technology (IJREAT), Volume 5, Issue 1, Feb - Mar, 2017, ISSN: 2320 – 8791.
- [10] Jaideep Srivastava, Robert Cooleyz, 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.
- [11] Lya Hulliyyatus Suadaa, "A Survey on Web Usage Mining Techniques and Applications", International Conference on Information Technology Systems and Innovations (ICITSI), Bandung-Bali, 24-27 November, 2104, pp. 39-43, ISBN: 978-1-4799-6526-7.
- [12] S. Prakasam, Prof. Dr. R. M. Suresh, "An agent-based Intelligent System to enhance E-learning through Mining Techniques", International Journal on Computer Science and Engineering (IJCSE), Vol. 02, No. 03, 2010, pp. 759-763, ISSN: 0975-3397.
- [13] P. Veeramuthu, Dr. R. Periyasamy, V. Sugasini, "Analysis of Student Result Using Clustering Techniques", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5(4), 2014, pp. 5092-5094, ISSN: 0975-9646.
- [14] Xue Sun, Wei Zhao, "Design and Implementation of an E-learning Model Based on WUM Technique", International Conference on E-Learning, E-Business, Enterprise Information Systems and E-Government, IEEE 2009, doi:10.1109/IEEE.2009.88, pp. 248-251.
- [15] 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.
- [16] 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, SuperStr Virgo, Singapore, pp. 1033-1036.
- [17] R. Cão, C.L de Freitas, "Modeling Critical Factors of Quality in e-Learning", 12th IEEE International Conference on Advanced Learning Technologies, 2012 pp. 422-426.
- [18] Rana Khudhair Abbas Ahmad, "COMBI: A New Combination of Web Mining Techniques for E-learning Personalization", International Journal of Technical Research and Applications, Volume 3, Issue 1, Jan-Feb 2015, pp. 104-108, e-ISSN: 2320-8163.