Recommender System Based On Web Usage Mining For Restructuring Of E-Learning Websites And Blogs

Sunil[1], Prof. M. N. Doja[2]

[1]Computer Engineering, Department of University Polytechnic, 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

Abstract

Nowadays, World Wide Web has become a popular medium to extract relevant and useful information from the rapidly growing huge amount of information particularly for education and e-learning. Web based learning environments are now days becoming are very popular and attractive medium for the learners. In a very short span of time it is observed a magnificent growth in e-learning also it plays an important role in higher education. E-learning system always provides more friendly and effective ways for students to increase their knowledge and improve the quality of education. Day by day a large number of e-learning websites and blogs are being designed to provide useful information to the learners. A well designed website is always considered as a successful website. It must be designed such a way that it can always satisfy the needs of the users. An optimized, updated and structured website always attracts learners. In an e-learning context a recommender system is a software agent that tries to intelligently recommend actions to a learner based on the actions of the previous learners. In this regard this paper discusses about the improved mining strategies and discusses about the available recommender agents that are required for maintaining and restructuring websites which in turn is helpful for developers to increase the visits of current learners and also attracts new learners.

Keywords— E-Learning, web mining, web usage mining, web widget.

1. Introduction

To support the learning system and to achieve its goals E-Learning is becoming an important tool. In the 1990's E-Learning became hot topic after the spread of the Internet. Although it has a relative short history, it is becoming an important part of the learning [1]. Some kinds of e-learning is always adopted by the majority of the universities within its learning system. E-Learning may either be synchronous or asynchronous.

Synchronous learning occurs in real-time, where all participants interacting at the same time. The exchange of ideas and information is done during the same period of time with one or more participants in synchronous learning. Some of the examples are online real-time live teacher instruction and feedback, face-to-face discussion, skype conversations, and chat rooms or virtual classrooms where everyone is online and working collaboratively at the same time. On the other hand Asynchronous learning is self-paced and allows participants to engage in the exchange of information or ideas without the dependency of other participants′ involvement at the same time. Asynchronous learning may use technologies such as blogs, discussion boards, email, and wikis as well as social networking using web 2.0, hypertext documents, web-supported textbooks, and audio video courses.

This paper discusses about the improved mining strategies and discusses about the available recommender agents that are required for maintaining and restructuring websites and blogs. Here we collected data of learners visit for the blog named 'History of the Computers' where a lot of information is present about the evolution of the computers that is useful for the learners. Learners visit the blog for desired information and their visiting patterns can be used to improve the blog structure, and then finally provide a better characteristic information service to the students/learners. Here a client side browser based application called web widget is used, which display data coming from different sources for blog and create web logs.

2. Related Work

Web usage mining has been one of the best areas of researches now days. Web usage mining technique being used widely to discover the user navigation patterns from web server logs.

Husna Sarirah Husin et al. [2] focus on news recommendation based on web usage and web content mining. In which they used web usage and web content mining techniques to recommend news articles to users, and by using web server logs from a Malaysian newspaper, Berita Harian and combined with the web content pages and discovered the web users navigation patterns.

Mirghani A. Eltahir et al. [3] proposed a technique for extracting knowledge from web server logs using web usage mining. In which they analyse the web logs of a website www.interactivegt.com with the help of deep log analyser program and by using web usage mining they discovered and understand the customers profile and their performance in terms of strength and weaknesses of their website.

Péter Tóth [4] proposed his work for adaptive online learning environment and web usage mining. There by the application of web mining method he was trying 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.

Wasvand Chandrama et al. [5] proposed a survey on data preprocessing method of web usage mining. They 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.

Ashika Gupta et al. [6] proposed their work on web usage mining using improved frequent pattern tree algorithms. Their research work focuses on web uses mining and specifically keeps tabs on running across the web utilization examples of sites from the server log records. The bonding of memory and time usage is compared by means of Apriory algorithm and improved frequent pattern tree algorithm.

A. Yadav et al. [17] proposed the effectiveness of the web usage mining techniques that can make e-learning environment more effective and 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 the learners.

3. Source of Data for Web Usage Mining

The data is the basics of a knowledge discovery process. As it is shown in Fig. 1 there are several possible data sources for the Web Usage Mining process. Each type has its own advantages and a little different focus.
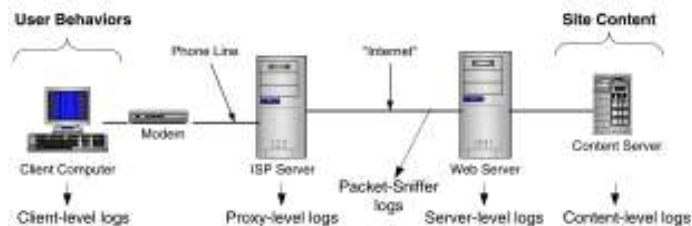

Fig. 1. Data Sources

Data Sources: There are three data sources, they are:

3.1 Server Level Collection

Because of their easy availability server level data collections are often used – every web server saves logs. Web server logs are the best known data source for web usage mining. 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.

3.2 Client Level Collection

Client level data collections are focused on tracking single user sessions across single or multiple web sites. 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's movement. How much time the user spends on a Web page it is recorded by the browser, if the visitors pushes the back or reload button, and many other valuable variables.

3.3 Proxy Level Collection

The last data source is proxy server logs. These are suitable for characterizing the browse behavior of a group of users sharing the same proxy server logs and save multiuser or multisite communication.

4. 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 [7]. Usage data captures the identity or origin of web users along with their browsing behavior at a web site. Fig. 2 shows web usage mining process. The Web usage mining is parsed into these distinctive phases or stages:
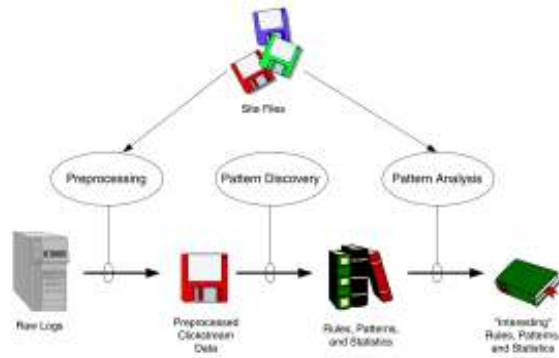
Fig. 2. Web Usage Mining Process

4.1 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 [8]. Fig. 3 shows usage preprocessing details.
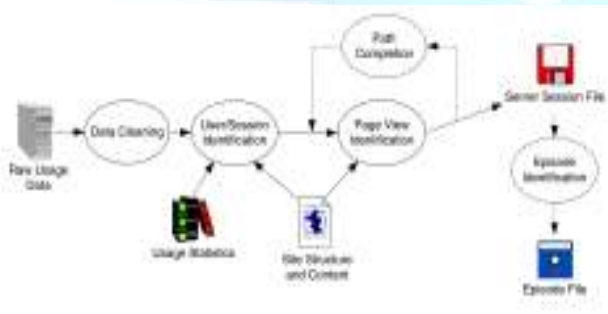

Fig. 3. Usage Preprocessing Details

4.2 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 1og.

4.3 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 [9].

5. Results

The proposed work analyses the web logs of the Blog- History of the Computers (http://www.history-of-the-computers.blogspot .in) and the access patterns of learners is collected with the help of recommender agents also called web widgets named flag counter and feedjit. The log files consisted of about 7 years data for the visiting patterns of the users from October' 2010 to January' 2017. The analyzed results can be seen as knowledge to answer these questions. What is the optimal structure and content of a web site in order to attract the maximum interest of visitors? User interacts with web by providing what set of keywords in their query.

The outcomes of these questions can be used for optimization, update and restructuring the blog to achieve the ultimate goal as to efficiently provide the useful information to the learners and also to attract the new learners. As Fig. 4. illustrates the GUI of the blog 'History of the Computers'. In this we are using a client side, browser based application called web widget, which display data coming from different sources for our own blog and create web logs.

Fig. 4. Blog-'History of the Computers'

A widget has the role of a transient or auxiliary application that means it just occupies a portion of a webpage and does something useful with information fetched from other websites and displayed in place and that will be combined with the web content pages to discover the web users' navigational patterns. Typically, web widgets are user interfaces and they are portable, full-fledged, self-contained, and mostly client-side application, providing a minimal set of functionality with considerably less complex and comprehensive. Widgets are expected to be re-usable, which is achieved by enabling widgets to be embedded in different platforms satisfying certain standards and specifications (e.g., W3C widget specifications). Although various technologies can be used to implement widgets (notably HTML and JavaScript, Java Applets, Flash etc.), cross-platform and device support is crucial due to re-usability considerations.

Data of web widget (feedjit) is shown in the Fig. 5; where web users' navigational patterns are given. The various fields appear are as: Country name of visitor, Browser, Operating system, web site used, time of visit. Many different types of results occur according to Web Usage Mining technique.

The (visual) knowledge can be easily understandable and usable by humans. Implementation of Web Usage Mining steps displays an example results below. Fig. 6 and 7 shows website popularity and accessed resources to know different users, Fig. 8 shows referring URLs to the blog and referring sites, referring search engines to the blog can be seen in the Fig. 9. Top search keywords for blog contents used by visitors are shown in the Fig. 10.

From the given users' navigational patterns in the Fig. 10, for the top search keywords from the blog contents it comes as a result that for a preferred keywords, the visitors are visiting the blog. As from the result it shows that for the keyword 'Astrolabe', maximum key searches has come, as about 6364 persons visited for its information through search engines. So from the result of users visiting pattern and their interest it shows that the blog can be modified on the basis of the visitors' requirement and more information can be added regarding 'Astrolabe'. Following are some examples of results divided according to user information registered in log files:

Fig. 5. Web log detail from web widget 'Feedjit'.

	Country	Visitors	Last New Visitor
1.	United States	5,743	January 19, 2017
2.	Philippines	1,461	December 4, 2016
3.	India	1,109	4 minutes ago
4.	France	825	January 22, 2017
5.	Canada	555	July 29, 2016
6.	United Kingdom	470	November 20, 2016
7.	Australia	220	February 22, 2016
8.	Spain	208	September 27, 2016
9.	Pakistan	175	October 7, 2015
10.	Brazil	156	December 17, 2016
11.	Mexico	138	January 9, 2017
12.	Jamaica	130	September 25, 2016
13.	United Arab Emirates	100	March 8, 2015
14.	Ireland	95	May 3, 2016
15.	Sri Lanka	92	January 3, 2017
16.	Germany	89	November 30, 2016
17.	South Africa	87	March 4, 2015
18.	Netherlands	80	December 18, 2016
19.	Turkey	68	August 9, 2016
20.	Switzerland	67	March 1, 2016
21.	Thailand	67	January 13, 2015
22.	Belgium	55	February 27, 2016
23.	Nigeria	53	October 25, 2016
24.	Indonesia	53	August 27, 2016
25.	Malaysia	52	October 28, 2015
26.	Poland	48	May 28, 2013
27.	Russia	43	November 15, 2016
28.	South Korea	40	February 2, 2015
29.	Japan	39	November 14, 2016
30.	Colombia	39	March 2, 2015
31.	Romania	38	November 9, 2014
32.	Israel	38	July 9, 2013
33.	New Zealand	37	March 11, 2015

Fig. 6. Top 33 Country visitors of the Blog.
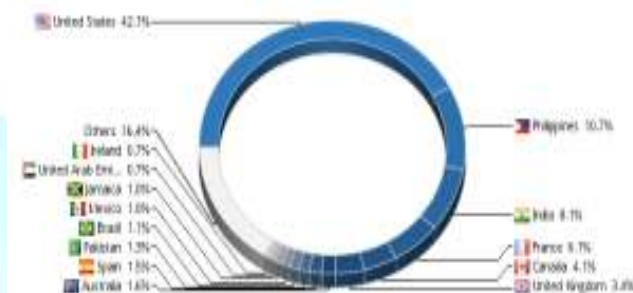
Fig. 7. Overview of visitors of the Blog.

Referring URLs

Entry	Pageviews
http://twiends.com/ViewSite.aspx?...	242
http://sundos.blogspot.in/	160
http://www.google.com/imgres?mu...	132
http://sundos.blogspot.com/	110
https://www.google.com.ph/	55
http://www.google.com/imgres?q=...	50
http://www.twiends.com/ViewSite a...	49
https://www.google.co.in/	41
http://twiends.com/visitsite.aspx?vi...	38
http://www.bing.com/images/searc...	29

Fig. 8. Referring URLs to the Blog.

Referring Sites

Entry	Pageviews
www.google.com	8808
www.google.com.ph	2513
www.google.co.in	1313
www.google.fr	1100
www.google.ca	677
www.google.co.uk	676
www.bing.com	372
twiends.com	281
www.google.es	214
www.google.com.au	210

Fig. 9. Referring Sites to the Blog.

Fig. 10. Top search keywords for blog contents.

6. Conclusion

E-learning is becoming an important tool to support the learning system and offers new possibilities in learning. Today's learners are highly dependent on web. As a large number of e-learning websites and blogs are being designed to provide useful information to the learners. In order to make a website popular among its visitors, website administrator and web designer should try to increase its effectiveness by understanding users and learning about them. The web designer must study the learners' requirements and their preference so as to maintain and restructure their websites and blogs. Web mining is successful research area evolving now days because it has impacted various application areas including data mining and its conceptualization with e-learning. Web usage mining is the procedure of finding out which users are looking for the internet and for what content. The proposed mining strategies and recommender agents may further be improved by employing innovative research areas like cloud computing, soft computing techniques.

References

[1] Dr. Jamil A. Itmazi, "E-Learning Systems and Tools" 1st editing, Publisher: Phillips Publishing, on April 2010, ISBN: 978-0-557-44966-8.

[2] Husna Sarirah Husin, James A. Thom and Xiuzhen Zhang, "News Recommendation Based on Web Usage and Web Content Mining", ICDE Workshop, 2013, pp. 326-329.

[3] Mirghani A. Eltahir and Anour F. A. Dafa-Alla, "Extracting Knowledge from Web Server Logs Using Web Usage Mining", International Conference on Computing, Electrical and Electronics Engineering (ICCEEE), 2013, pp. 413-417.

[4] Péter Tóth, "Adaptive Online Learning Environment and Web Usage Mining", 8th IEEE International Symposium on Applied Computational Intelligence and Infomatics, Timisoara, Romania, May 23-25, 2013, pp. 61-66.

[5] 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.

[6] Ashika Gupta, Rakhi Arora, Ranjana Sikarwar and Neha Saxena, "Web Usage Mining Using Improved Frequent Pattern Tree Algorithms", International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014, pp. 573-578.

[7] 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.

[8] Arvind Kumar Dangi and Sunita Sangwan, "A new Approach for User Identification in Web Usage Mining Preprocessing", IOSR-JCE, May 2013.

[9] 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.

[10] B. Naveena Devi, O. Sreevani, "Dynamic Modelling Approach for Web Usage Mining Using Open Web Resources", International Journal of Engineering Science and Technology (IJEST), Vol. 2(10), 2010, pp. 5605-5610. ISSN: 0975-5462 (2010).

[11] Mohd. Majid Shaikh, Dr. V. M. Patil, Prof. S. V. Dhopte, "Web Mining: Application for Web Based Distance Education Platform", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (3), 2014, pp. 4424-4425.

[12] B. Santosh Kumar, K. V. Rukmani, " Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms", International Journal of Advanve Networking and Applications, Vol. 01, Issue: 06, pp 400-406 (2010).

[13] C. J. Carmona, S. Ramirez-Gallego, F. Torres, E. Bernal, M. J. Del Jesus, S, Garcia, " Web Usage Mining to improve the Design of an E-Commerce Website: OrOliveSur.com", Intenational Journal of Expert Systems with Applcations", 2012 Elsevier.

[14] Olatz Arbelaitz, Ibai Gurrutxage, Aizea Lojo, Lavier Muguerza, Jesus Maria Perez, Inigo Perona, " Web Usage

and Content Mining to Extract Knowledge for modeling the Users of the Bidasoa Turismo Website and to adapt it.", Intenational Journal of Expert Systems with Applcations", 2013 Elsevier.

[15] 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 (3), 2014, pp. 5092-5094.

[16] Vishal Venkat Raman, Swapnil Vijay, Sharmila Banu K.' " Identifying Customer Interest in Real Estate Using Data Mining Techniques", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (3), 2014, pp. 3081-3084.

[17] Anuradha Yadav, Satbir Jain, "Analyses of Web Usage Mining Techniques to enhance to Capabilities of E-learning Environment", IEEE, pp. 223-225.