International Journal of Information Technology and Knowledge Management
July-December 2010, Volume 3, No. 2, pp. 729-732

AN ENTRENCHED INFORMATION ORGANIZATION IN WEB-BASE EDIFICATION

Mohd. Husain[1], Raj Gaurang Tiwari[2] & Bineet Gupta[3]

Modern distance education essentially differs from traditional pattern of education in providing personalized education support service. Application of data mining techniques to the World Wide Web, referred to as Web mining, which has been the focus of several recent research projects and papers. One of several possibilities can be its application to the distance education. Taken as a whole, the emerging trends in distance education are facilitating its usability on the Internet. With the explosive growth of information sources available on the WWW, it has increasingly understood for Web mining to be suitable for keeping pace with the trends in education, such as mass customization. In this paper, we define Web mining and present an overview of distance education. We describe the possibilities of application of Web mining to distance education, and, consequently, show that the use of Web mining for educational purposes is of great interest.
Keywords: Internet, Web Mining, Distance Education, Path Analysis, Association Rules

1. Introduction

In the age of the Internet, the digital technology is being adapted to humans in education through four ways [6]. The first trend is a better understanding of mental processes and structures. With this knowledge, the technology can be better adapted to humans. Another trend is a shift away from mass standardization of education toward mass customization - customizing both education and technology to individual learner needs and preferences. The third trend is a greater appreciation of the importance of context for comprehension and learning. Another trend is that the knowledge base is expanding rapidly, much faster than ever before. Among them, mass customization pays attention to adapting instruction to individual learners, which is a shift from mass standardization - adapting learners to the system. Web mining can be helpful in pursuing mass customization in education. In addition, it can be helpful in identifying virtual knowledge structure in a Web-based education. Web mining can be broadly defined as the discovery and analysis of useful information from the World Wide Web [2]. This describes the automatic search of information resources available on-line, i.e., Web content mining, and the discovery of user access patterns from Web servers, i.e., Web usage mining. In this paper, we provide a framework associated with mass customization and virtual knowledge structure in a Web-base education, to which Web usage mining can mainly devote.

[1]Department of Computer Sc. & Engineering, Azad I.E.T., Lucknow, India

[2]Department of Computer Application, Azad I.E.T., Lucknow, India

[3]Department of Computer Science, Mizan Tepi University, ETHIOPIA

Email: mohd.husain90@gmail.com, rajgaurang@gmail.com, bineet777@gmail.com

2. Education and Web Mining

2.1. Education

During the past several decades, proclaimed revolutions in education have hardly proved to produce substantial changes in educational practice. Despite numerous initiatives to innovate educational systems and to improve the overall quality of educational approaches, little seems to have changed. The most common concepts and parameters of tutoring, classrooms, instruction, homework and examinations seem to be barely touched. And, educational innovation is easily confused with the mere introduction of new technologies [8]. The four major factors [9] which have been identified to clear the way for substantial innovations are:

1. Convergence of classroom teaching and distance learning: Classroom teaching, which is an essential condition for face-to-face education, is composed of classroom, the physical presence of some students and a teacher. In distance education, however, students work individually at their own homes using a set of self-contained study materials, i.e., computer program and the Internet. Time and space limitation of education disappears. In addition, individualized educational program and self-paced instruction enables students to maximize their ability.
2. New Avenue for collaborative learning: From an end-user's perspective, software becomes simpler and many programs can be executes without any training or without any knowledge of specific commands. Especially, the usefulness of computers is largely enhanced by telecommunication facilities, in an asynchronous mode like e-mail or news, or in

a synchronous mode, like videoconferencing, Groupware, a whole new class of software application, supports all kind of collaborative works. Clearly, such new technological tools introduce new ways of collaborative learning that necessarily presuppose the use of computers.

3. Changing student-tutor relationship: The worldwide computer network is assumed to have a tremendous impact on existing social and cultural patterns including educational patterns. As a consequence, some basic suppositions of educational systems are affected. Most important thing is the change of the role and position of a tutor. Tutor takes the role of coach who communicates and leads the course of their learning path. So, student will be grouped, will learn from each other, and occasionally will assess each other.

4. Changing basic principle: Constructivism: Having recognized these changes in educational systems, many educators have adopted constructivism as an alternative learning paradigm. According to constructivism, learners are not passive absorbers of external inputs, but actively construct meaningful knowledge from individual experiences.

2.2. Web Mining

Web mining has been used in two distinct ways. The first, called Web content mining is the process of information discovery from sources across the WWW. In recent years it has prompted researchers to develop more intelligent tools for information retrieval, such as intelligent Web agents, and to extend data mining techniques to provide a higher level of organization for semi-structured data available on the Web [1]. The second, called Web usage mining, is the process of the automatic discovery of user browsing and access patterns from Web servers [2]. Organizations, which run distance education sites, collect large volumes of data, generated automatically by Web servers and collected in server access logs. Other sources of user information include referrer logs that contain information about the referring pages for each page reference, and user registration or survey data gathered via CGI scripts. Analyzing such data can help organizations determine the thinking styles of learners, cross studying patterns across subjects, and effectiveness of a web site structure. It can also provide information on how to restructure a Web site to create a more effective Web site presence, and shed light on more effective management of collaborative study group communication and Web server infrastructure.

3. FRAME WORK

We have developed a framework for Web usage mining in distance education, which is presented in Figure 1.

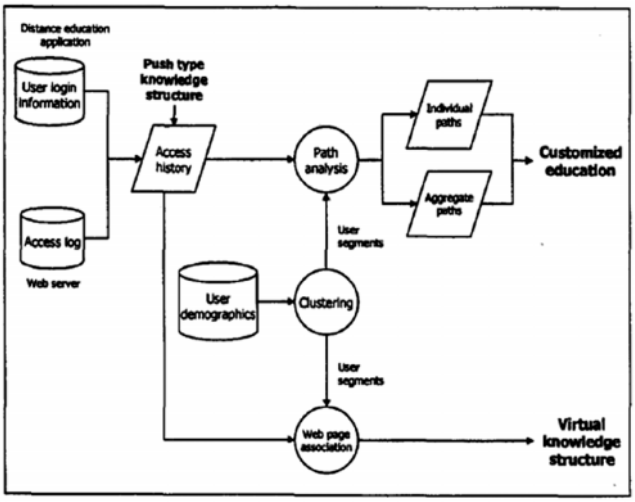

Fig. 1: Web Mining for Distance Education

The framework starts with using Web server logs and user login information, which are stored, maintained within a distance education system. Because the log cannot solely carry on the complete information for Web mining and, therefore, include the difficulty in identification of unique learners as well as learner sessions or transactions, it is necessary to develop a framework to integrate any Web information with back-end operational data, such as user login information [11]. Everyone who has a distance education has to register at the system and his/her demographic information is kept well in a user demographics database.

Before extracting access histories, on which the mining algorithms can be run, of a learner, a number of data preprocessing issues, such as data cleaning and transaction identification, have to be addressed.

The major preprocessing task is data cleaning. Techniques to clean a server log to eliminate irrelevant items are of importance for this type of Web log analysis. Elimination of irrelevant items can be reasonably accomplished by checking the suffix of a file name in the URL (Uniform Resource Locator). We can remove all log entries with filename suffixes such as gif, jpeg, GIF, PEG, which indicate graphic files. Identifying individual learners and their sessions can be done relatively easily because the system keeps login histories of each learner.

It is clear that extracted access histories of each individual learner are representing the physical layout of Web sites, with Web pages and hypertext links between pages, which are just given by distance educators. Once user access histories have been identified, there are several kinds of access pattern mining, such as path analysis, discovery of association rules and sequential patterns, and clustering and classification. We perform Web page traversal path analysis for customized education, and Web page associations for virtual knowledge structures, which could

be formed by learners themselves as they navigate Web pages.

3.1. Customized Education

There are many different types of path analyses for customizing education. The most obvious is a graph representation [13], since a graph represents some relation defined on Web pages; a graph with Web pages as nodes and hypertext links between pages as directed edges, a graph with edges representing similarity between pages, or creating edges that give the number of users that go from one page to another. Our work relating with path analysis involves determining most frequently traversal patterns from the physical layout of a Web site. The path analysis is performed from the two points of view: aggregate and individual paths. The first part, aggregate paths, includes the process of clustering registered learners[1]. A registration at the onset of learning can capture important personal information (gender, age, ZIP code, and education level) that can be later on enhanced and mined A Web site database, which is especially created for user registration, can be segmented by one of clustering techniques such as a self-organizing map, also known as a Kohonen feature map, to discover learners with similar characteristics. The resulting learner segments with access histories can be used to determine most frequently visited paths of learners in each segment in a Web site (aggregate Examples of aggregate paths that can be discovered through path analysis are listed as the following Table 1 :

Table 1
Aggregate Paths of Learners

Count	Aggregate paths	Segment
52%	/InformationlDataMining/ Classification->Information/ ArtificialIntelligence/ NeuralNetwork	High educational levellearners
65%	/Information/DataStructure/ Tree- >/Information/ DataStructure/StackQueue ->/Information/ DataStructure/List	Low educational levellearners

The first aggregate path indicates 52% of learners in a segment with high educational level start at "/Information/ DataMining/Classification", and proceeds a more advanced topic, such as "/Information/ArtificialIntelligence/ NeuralNetwork". The second aggregate path states that 65% of learners with low education, at first, study on a data structure of tree in a subject, "Information". But they go back to the data structures of stack and queue, and list for acquiring background knowledge to better understand the tree structure.

The second part of path analysis is about discovering individual paths. It amounts to determine a set of frequently visited Web pages accessed by a learner in his/her visits to the server during a certain period of time. Discovery of such aggregate and individual paths for learners engaged in distance education can help in the development of effective customized education. In addition, aggregate paths discovered from WWW access logs can give an indication of how to best organize the educator organization's Web space. The aggregate paths make suggestions on learning sequences to learners who belong to the same segment and share similar characteristics (e.g., education level) [3]. Later on, they can facilitate the development and execution of future Web space design, such as dynamically changing a particular page (contents, hypertext links) for learners belonging in different segments.

3.2. Virtual Knowledge Structure

Discovering association rules is to find all associations and correlations among Web pages where the presence of one set of Web pages in a transaction implies (with a certain degree of confidence and support) the presence of other pages. In doing so, we discover the correlations among references to various Web pages available The first rule in the Table 2 shows that 45% of learners who belong to a high educational level segment and access "/Information/ Data Structure", also access "/Mathematics/Algorithm". The second rule states that 53% of learners in low education who access "/Mathematics/Derivative", make reference to "/Help/ Mathematics". The last rule indicates that 34% of all learners go around "/Bulletin" and "/Chatting" regardless of their educational levels. The discovery of association rules in Web server access logs allows Web-based distance educators to identify virtual knowledge structures against push type knowledge structures and helps in reorganizing Web space based on these structures. We call it "virtual knowledge structure" because it includes differences between the physical topology of Web spaces and user access patterns [4].

Table 2
Association Rule for Discovering Virtual Knowledge Structure

Correlation	Association rules	Segment
45%	(/Information/ DataStructure, Mathematics/ Algorithm)	High Educational Level Learners
53%	(MathematicsDerivative, /Help/Mathematics)	Low Educational Level Learners
34%	(Bulletin/Chatting)	All Learners

4. CONCLUSION

In this paper, we presented the applicability of Web mining, especially Web usage mining, for enhancement of distance education. It could improve the effectiveness and efficiency of distance education along two dimensions. The first dimension was that how mass customizing education was achieved through Web mining. Discovery of aggregate and individual paths for learners engaged in distance education could help in the development of effective customized education, and give an indication of how to best organize the educator organization's Web space. The second dimension was that how virtual knowledge structure was identified via Web mining. The discovery of association rules could make it possible for Web-based distance educators to identify virtual knowledge structures and to recognize Web Space based on these structures.

REFERENCES

[1] Tiwari Raj Gaurang, Singh Hemant Kumar "Significance of State Cloning Concept in Clustering Clickstreasm" in the Proceedings of National Conference on Emerging Technologies in Computer Science (ETCS-2007) MIET, Meerut, India, Sep 23, 2007, pp 227-233

[2] Tiwari Raj Gaurang, Singh Hemant Kumar "Mining Interesting Knowledge and Pattern Discovery from Weblogs" in National Seminar Held at IIET, Bareilly, August 2007.

[3] C. YanGu, "A Study on the Suggesting Learning Path of E-learning", in Institute of Information Management, Master Yunlin, Taiwan: National Yunlin University of Science & Technology, 2005.

[4] G. Piccoli, R. Ahmad, and B. Ives, "Web-based Virtual Learning Environments: a Research Framework and a Preliminary Assessment of Effectiveness in Basic IT Skills Training", MIS Quarterly, 25, pp. 401-426, 2001.

[5] S. S. Liaw, "Considerations for Developing Constructivist Web-based Learning", International Journal of Instructional Media, 31, pp. 309-321, 2004.

[6] Hilary McLellan, 'Being Digital': Implications for Education, Educational Technology, Nov-Dec, 1996.

[7] R. Cooley, B. Mobasher, J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", IEEE, 1997.

[8] Kaufman, R., "The Internet as the Ultimate Technology and Panacea", Educational Technology, 38(1), 63-63, 1998.

[9] Westera, Wim, Paradoxes in Open, "Networked Learning Environments: Toward a Paradigm Shift", Educational Technology, Jan-Feb 1999, 17-23, 1999.

[10] Julie McNamara, "Cleaning Up Web Data", DB2 Magazine, Qtr 4, 1999, Online Edition.

[11] M. S. Chen, J. S. Park, and P. S. Yu, "Data Mining for Path Traversal Patterns in a Web Environment", Proceedings of the 16th International Conference on Distributed Computing Systems, 1996.

[12] R. Cooley, B. Mobasher, J. Srivastava, "Grouping Web Page References into Transactions for Mining World Wide Web Browsing Patterns", Technical Report TR 97- 021, University of Minnesota, Dept. of Computer Science, Minneapolis, 1997

[13] P. Pirolli, J. Pitkow, R. Rao, "Silk Fiom a Sow's Ear: Extracting usable Structures from the Web", Proceedings of 1996 Conference on Human Factors in Computing Systems (CHI- 96), Vancouver, British Columbia, Canada, 1996.