Determining the Web Mining Techniques in E-Learning Domain

Punam
Asst. Professor Deptt. Of Computer Science
University College Miranpur, Patiala, India

Abstract: One of the key components of on-line learning at present is the course's web site, which is considered as a virtual lecture room of the course. It provides access to dynamic learning materials, and facilitate students to interface with the learning materials, with each other and with the teaching personnel. As on-line teaching requires reasonable contribution of intellectual and financial efforts, it is enormously vital to evaluate its academic effectiveness. The paper illustrates the use of Web mining as a mode to get better on line instruction. This paper exhibits the usage and content mining techniques as an objective for evaluating student's behavior in an on-line learning environment. Nonetheless, we argue that information thus gained can be of use to improve on-line teaching.

Keywords: E-Learning, Web Mining, Education Data Mining (EDM), World Wide Web.

I.   INTRODUCTION

Today, predicting the direction of alumni and students is one of the most important challenge that faces by our higher education. Institutions would want to know, for instance, which students will join up specific courses, and which students will require help in order to graduate. Additionally to current challenge, classical problems like enrollment management time-to-degree still inspires institutions of higher education to look for improved solutions.

With the growth of information sources available on www, it becomes necessary for educational institutions to use automated tools to discover required resources of information, record and analyze student's patterns. The best way to gather and analyze these student and organization challenges is through web mining. These factors make the requirement of creating client and server side intelligent systems which will productively mine for knowledge. Web mining has defined as to improve the power of web search engine by identifying the favorable information from www.

Web mining could benefit E-learning like following:

1) *Comprehend learner attitude*
University instructors and administers may be able to benefit the usage of E-learning systems by understanding the dynamic attitude of students in the web systems.

2) M*easure performance of E-learning system*
Behavior patterns might be aligned with performance of system and allow more customized system configuration.

Instructors and administrators might have the option to find raised and squat use regions of the E-learning system and change sources to upgrade the system's technical performance.

3) *Measure the achievement of instructional efforts*

Students use email, feedback forms, web forum etc. to express their issues and ask questions in E-learning system. This information are totally recorded in E-learning system. Web mining provide help in quantitative feedback about the results of instructor's activity.

II. WEB USAGE MINING

Web server's stores data of various user interactions of received request of resources. Examining the logs of web access of different websites can help to recognize web structure and behavior of user, which helps in design improvement of this vast collection of sources. Web mining is the process toward identifying what users are finding on web. Few users might be taking multimedia data whereas some other taking textual data. It helps in discovering the pattern of search for a specific group of users relating to a specific area. Web usage mining is used to find usage patterns from Web information, so as to comprehend and better serve the requirements of Web-based applications.

Depending on its kind of usage, web usage mining can be classified in three categories:

- **Application Server Data**: Application servers of commercial use have vital aspects to permit online business applications to be based over them with small effort. The main factor is the capability to follow different kinds of business events and log them in application server logs.
- **Web Server Data**: The user logs like IP addresses, access time and page reference are gathered by web server.
- **Application level data:** New types of cases can be described in an application, and logging can be turned on for them in this manner consequently producing histories of these specially defined events. Many end applications need a mix at least one of the strategies applied in the above categories.

There are two foremost inclinations in Web Usage Mining driven by the functions:

A. General Access Tracking of Patterns

To understand access trends and patterns, the web logs are analyzed by the general access pattern tracking. These analyses can reveal insight into better structure and gathering of suppliers of resource. By implementing data mining systems on web access logs discloses concerning access designs which can be utilized to rebuild websites in an increasingly proficient gathering, pinpoint viable advertising areas, and point to particular users for explicit selling advertisements.

B. Customized Usage Tracking

Individual trends analyzes are tracked by customized usage. To customize different websites to users is its objective. The information shows the website structure profundity and arrangement of sources can be dynamically custom made for every user after some time based patterns of their access. While it is interesting to see the different potential uses of web log record analysis, it is vital to understand that the success of these applications relies upon what and how much reliable and valid learning one can find from the enormous raw log data. Current web servers save confined data about their accesses. A few scripts especially custom-made for few websites might store extra data. In any case, for a successful web utilization mining, a significant cleaning and data change step before analysis might be required.

III. WEB CONTENT MINING

Web content mining is the procedure to find valuable information in the web from video, image, audio and text information . Web content mining also referred as web text mining, because the textual content material is the most extensively area of research. The technologies that are commonly utilized in web content mining are Natural language preparing (NLP) and Information retrieval (IR). Despite the fact that information mining is a generally new term, the technology isn't. Persistent developments in computer disk storage, processing power, and statistical software are drastically expanding precision of analysis while driving down the expense. Web content mining is an automated procedure that goes beyond keyword extraction. Since the content material of a text document offers no machine-readable semantic, a few methodologies have recommended to rebuild the document content in a representation that can be worked by machines.

In web content mining we look at the real content of site pages and afterward some knowledge discovery procedure is performed to find out the pages themselves and their connections. Normally this is carried to arrange out a gathering of documents into linked categories. It is particularly useful for web search engines, after all it permits users to rapidly discover data that they are searching for in contrast to common "endless" ranked list. When representing internet document and text for classification and grouping, a vector-space model is commonly utilized. In this model, every feasible term which can show up in a report turns into a feature dimension. Assigned value of every element of a record might indicate the number of times the representing term occur on it or it might be a weight that takes into knowledge other frequency information, for example, number of records where upon the terms occur. Techniques are yet to come which using lexicons for content interpretation. Web content mining strategies are of two types: which directly mine the content material of documents and which enhance the content material search of other tools like search engines.

IV. WEB STRUCTURE MINING

WWW uncover a lot of data just than the data enclosed in documents. For instance, connections coming into a document show record's demand, while connections which popping out of a document demonstrate extravagance and assortment of points canvassed in the report. At the point when a paper is referred to regularly, it should be significant. The Page Rank and smart strategies exploit this data sent by the links to discover appropriate pages on internet. By methods of counters, more high levels collect the number of artifacts by the ideas they hold. Hyperlinks counters, out and in reports, remember design of the web artifact outlined. Web structure mining is the way to analyze the web site's nodes and connection structure of a site by using graph theory. Web structure mining can be split into two kinds, according to the type of web structural data. The firs kind of web structure mining is to deriving patterns from hyperlinks in the web. A hyperlink is a structural aspect which associates the page to a substitute location. Mining the report structure will be the second type of the web structure mining. Tree-like structure is used to examine and depict the HTML tags inside the page of website.

Structure mining plans to derive formerly undiscovered relations between pages of web. This structure mining gives an association to link data of its own Web website to empower navigation and batch data into maps of website. With improved navigation of Web pages other educational Websites, associating the mentioned information to a web search engine results to be more efficient. This more strong connection permits creating traffic to website to provide outcomes that are more profitable. The extra links gave inside the relationship of the website pages allow the navigation to yield the link hierarchy allowing navigation ease. The enhanced navigation brings the alumni and students to exact locations giving the required data, demonstrating more useful in clicks to a specific website. In this way, structure mining utilization and web mining can give strategic results to advertising of a Web webpage for education. A lot of traffic coordinated to the Web pages of a particular website expands the level of return appearance to the webpage and recall by search engines describing to the data. This additionally empowers promoting strategies of educational organizations to give results that are a lot of beneficial through navigation of the pages linking to the home page of the site itself.

V. WEB MINING APPLICATIONS IN EDUCATION

In all fields of society, communication technologies and information had busted in educational system through web these days. Therefore it is feasible to trap and gather data of all sorts in a simple manner and minimal effort, for example, universities and high schools registers, computer-supported collaborative learning systems, administrative data of schools, e-learning platforms activity record, computerized students' academic records, and so on. Educational Data Mining (EDM) is an application of web mining. EDM is the way toward changing raw data in useful information arranged by educational systems that can be utilized to take right finding and answer the research queries. In educational field, EDM is the one of application of data mining, with the motive to get a better comprehension of learning process of students and their worldwide presence on it, conducted for quality improvement and cost-viability of the education system. There are three goals can be recognized in EDM: enhancement of the students' academic performance, pedagogic objectives which help in the plan of didactic contents, management objectives needed for maintenance of educational infrastructures and organization optimization, student interest area's, more required courses and commercial objectives which permits to make market segmentation and encourages the enrollment of students, that is particularly significant on account of private schooling. EDM is an emerging trend on the intersection of pedagogy and data mining. Pedagogy contributes with the inherent knowledge of process of learning on one hand. And data mining includes the analysis and data modelling methods on other hand. EDM is a technology that coordinates multidisciplinary techniques, which is used for retrieval of information through SQL queries from database with the utilization of algorithms of automatic learning coming from field of Artificial Intelligence. Institutions of higher education can utilize grouping for a complete analysis of students qualities and can utilize estimation to foresee the probability of an assortment of results, for example, transferability, retention, persistence, maintenance, and success of course.

VI. CONCLUSIONS

With the increasing popularity of the WWW, there is a developing need to strengthen techniques and tools that will assist to enhance its standard usefulness. Web mining can act as an effective analytical tool that allows educational organizations proactively manage results of students, to better allocate staff and resources and enhance the effectiveness of alumni development. Since one of the key objective of the Web is to act as an overall disbursed information source, a wide variety of efforts are underway to develop techniques that will make it extra beneficial in this manner. Utilization of Web mining has been refer to various types of techniques which incorporate a broad range of issues. We gave a point by point overview of the efforts in this area, despite the survey is short due to the fact of the novelty of area. We gave a general survey of a system to do Web mining, and diagnose few of the problems and issues in this area that might also need further research and development.

VII. REFERENCES

[1] Nasraoui O., Petenes C., "Combining Web Usage Mining and Fuzzy Inference for Website Personalization", in Proc. of WebKDD 2003 – KDD Workshop on Web mining as a Premise to Effective and Intelligent Web Applications, Washington DC, August 2003, p. 37

[2] H. Chen and M. Chau, "Web mining: Machine learning for Web applications," Anu. Rev. Inf. Sci., vol. 38, pp. 289–329, 2004.

[3] M. Chau, B. Shiu, I. Chan, and H. Chen, "Redips: Backlink search and analysis on the Web for business intelligence analysis," J. Amer. Soc.Inf. Sci. Technol., vol. 58, no. 3, pp. 351–365, 2007.

[4] X. Fang, M. Chau, P. J. Hu, Z. Yang, and O. R. L. Sheng, "Web mining based objective metrics for measuring Website navigability," in Proc. Int. Conf. Inf. Syst., Milwaukee, WI, 2006.

[5] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web," in Proc. 2005 WWW Conf., Chiba, Japan, 2005.

[6] R. Kosala and H. Blockeel, "Web mining research: A survey," ACM SIGKDD Explor., vol. 2, no. 1, pp. 1–15, 2000.

[7] J. Cho, H. Garcia-Molina, and L. Page, "Efficient crawling through URL ordering," in Proc. 1998 WWW Conf., Brisbane, Australia, 1998.

[8] S. Chakrabarti, M. V. D. Berg, and B. Dom, "Focused crawling: A new approach to topic-specificWeb resource discovery," in Proc. 1999 WWW Conf., Toronto, Canada, 1999.

[9] M. Chau and H. Chen, "Comparison of three vertical search spiders," IEEE Computer, vol. 36, no. 5, pp. 56–62, 2003.

[10] O. Zamir and O. Etzioni, "Grouper: A dynamic clustering interface to Web search results," in Proc. 1999 WWW Conf., Toronto, Canada,1999.

[11] W. Y. Chung, G. Lai, A. Bonillas, W. Xi, and H. Chen, "Organizing domain- specific information on the Web: An experiment on the Spanish business Web directory," Int. J. Hum.-Comput. St., vol. 66, no. 2, pp. 51–66, 2008.

[12] Lillian Clark, I-Hsien Ting, Chris Kimble, Peter Wright, Daniel Kudenko (2006)"Combining ethnographic and click stream data to identify user Web browsing strategies" Journal of Information Research, Vol. 11 No. 2, January 2006

[13] Eirinaki, M., Vazirgiannis, M. (2003) "Web Mining for Web Personalization", ACM Transactions on Internet Technology, Vol.3, No.1, February 2003.