Data Mining Techniques to Discover Students Visiting Patterns in E-learning Resources

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

[1]Section of Computer Engg., Faculty of Engg. & Technology, Jamia Millia Islamia, New Delhi, India

[2]Department of Computer Engg., Faculty of Engg. & Technology, Jamia Millia Islamia, New Delhi, India

[1] sunil@jmi.ac.in; [2] mdoja@jmi.ac.in

Abstract-- In recent times, the rapid progress of internet technology has triggered the extensive development of web-based learning environments in the educational world. Online learning resources provide various types of online learning assets like tutorials, e-books, scientific articles, etc. Nowadays students prefer E-learning resources for learning and collecting useful information through it. As students access the E-learning resources like websites or blogs, the sequences of their access containing the web pages are stored in the web server logs. These web server logs can be analyzed to find the students visiting patterns. In this paper a set of tools that are based on data mining techniques such as clustering, and association mining, combined with collaborative filtering techniques are discussed. Then some methods are presented to analyze the browsing web log data to construct a browsing behavioural model that is helpful in supporting E-learning resources.

Keywords— Data Mining, web usage mining, E-learning, recommendation system, visiting patterns.

I. INTRODUCTION

Within a decade, the World Wide Web has become a persistent medium that has changed entirely, and possibly permanently, the way information and knowledge are transmitted and shared throughout the world. The learning culture has not partial itself to the role of submissive player in this recitation legend, but it has been at the front position of most of the changes [1]. E-learning, a new context for education where large amounts of information relating the variety of the teaching–learning interactions are continually generated and universally available. The fast developments in the communication and information technologies have increases the widespread use of internet, which has changed the learning behaviours of the students. The internet has become an information nucleus that is reachable in an insidious manner. It is a very excellent platform for conveying and gaining information for students. E-learning resources engage students in a process of learning by providing them required useful information promptly and in valuable way [2]. Data Mining is the process of extracting valuable information or knowledge from huge databases which could be used in decision making procedure for building valuable decisions [3]. Pattern mining is one of the most important areas of research in the field of data mining as patterns from a sequence database. Many educational institutions are fascinated in mining sequential patterns from their E-learning resources due to the huge amounts of data being collected and stored continuously through visiting of E-learning resources by the students. Introducing data mining to pertain on web-based learning systems is one of the imperative methods to understand the intelligent and individual E-learning environments [4]. The web based systems can analyze, categorize and cleanup the mining result sets and give

suitable schema for every student. Data mining techniques can extract the valuable information about students accesses obtained from the web server logs. By analyzing the extractions it can be supportive to construct an intelligent and individual E-learning resource for students.

II. DATA MINING

Data Mining is habitually defined as finding hidden information in a database. On the other hand, it has been called investigative data analysis; data focused discovery and deductive learning. Data mining is measured as the most appropriate technology suitable in giving supplementary handy into the students, teachers and other educational staff behaviour and performing as an energetic automated associate in serving them for making enhanced decisions on their educational activities [5]. Data mining technology can help bridging knowledge gaps in educational system. Therefore the hidden patterns, association and anomalies, which are discovered by data mining techniques, can be used to improve the usefulness, competence and the speed of the processes. As a result, this enhancement may convey a lot of advantages to the educational system such as make best use of educational system competence, declining student's drop-out rate, escalating student's promotion rate, rising student's withholding rate, growing student's success, rising student's learning outcome, and dropping the cost of system processes.

A. Association mining

Association mining is one of the largely fine studied methods in data mining [6]. It has hand out as a valuable tool for discovering linked items in a hefty transactional database. It produces if-then statements relating attribute values. An association rule $X \Rightarrow Y$ articulates that in those transactions in the database where X occurs; there is a high possibility of having Y as well. X and Y are called the antecedent and consequent of the rule. The potency of such a rule is calculated by its support and confidence. The support of the rule is the percentage of transactions in the database that restrain both the antecedent and the consequent. The confidence of the rule is the percentage of transactions with X in the database that restrain the consequent Y also. Association rule mining has been applied to E-learning resources for conventional association analysis such as building recommender agents for E-learning activities, involuntarily guiding the students' actions and intelligently generates and recommends learning materials to the students, finding out the relationships between each pattern of students' behaviour, identifying attributes distinguish patterns of performance inconsistency among diverse groups of students.

B. Clustering

Data mining permits the user to analyze data from diverse dimensions, categorize it and recapitulate the relationship recognize during mining process [5]. Data mining techniques are used to function on large volume of data to discover hidden pattern and association that are supportive in decision making. Different data mining techniques are used in the field of education. Cluster analysis is used to fragment a hefty set of data into subsets. Each cluster is a compilation of data objects that are analogous to another located within the same cluster but divergent to objects in other cluster. Clustering is one of the fundamental techniques frequently used in analyzing data sets. It is a technique to group collectively data items or users with similar characteristics. Clustering of user information or data items can assist the development and carrying out of potential marketing strategies [7]. The users who have related navigation patterns, group of those users can be exposed by clustering. Here, the cluster analysis is used to subdivision students into groups according to their characteristics. Clustering can be considered the most important unsupervised learning technique.

C. Web usage mining

Web usage mining is relevance of data mining techniques to determine interesting usage patterns from web usage data which is collected from the students visit for E-learning resources, in order to understand and better serve the needs of web based applications [8, 9]. Web usage data confine the identity or origin of web users along with their browsing behaviour from E-learning resources. As shown in the fig.1 web usage mining process, the web usage mining is parsed into these characteristic phases or stages they are data collection and preprocessing, pattern discovery and pattern analysis.
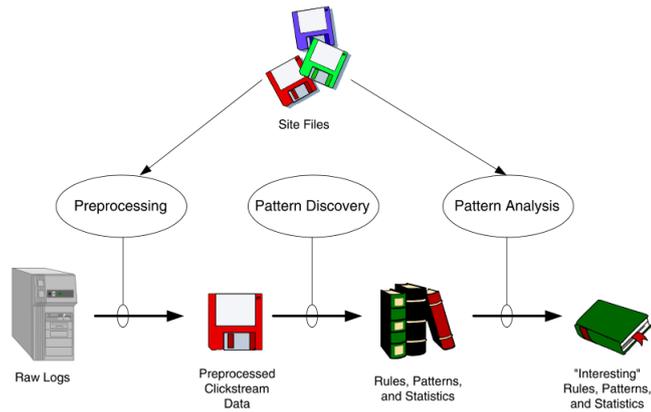
Fig.1 Web usage mining process

The revealed knowledge signifying students' navigational behaviour that is useful for the system to personalize the E-learning resources according to each student's behaviour and profile [10].

III.RECOMMENDATION SYSTEMS

Recommendation systems are proposed to wrap the space between information collection and analysis, by straining all the offered data, and presenting the most appropriate items to the students [11]. Currently the recommendation systems are broadly adopted in various fields of data mining for the recommendations. The E-learning resources can be recommended based on the top overall students visiting on an E-learning website, or on an analysis of the past learners visiting behaviour as a prediction for future visiting behaviour. The forms of recommendation include signifying useful E-learning resources to the students and providing personalized learning information [12]. Generally, these recommendation techniques are the ingredient of personalization on an E-learning website because they help the website to acclimatize itself to each student.
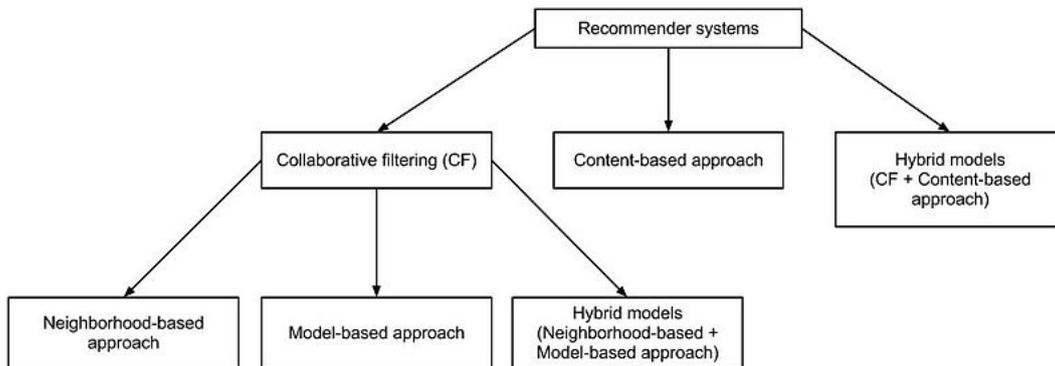

Fig.2 Recommender System

A. Types of Recommender Systems

1) Collaborative Filtering Recommendation Systems

In this approach of recommendation systems, a student is recommended learning resource that other students with similar tastes and preferences selected in the past. For example, many book recommendation systems assist learners in finding suitable information on the internet. There are two major approaches for collaborative filtering algorithms: They are Model-based approaches and Memory-based approaches.

2) Content-based filtering Recommendation systems

This approach uses a set of diverse features of an item to recommend more items with similar properties. Content-based filtering is based on item description and user preference. In this type of recommendation system, a student is recommended E-learning resources similar to the ones student preferred in the past. For example, if a student named 'Mr. Abc' reads an E-learning resource like many online research papers on the topic web data mining, then the content-based recommendation techniques will be able to recommend other web data mining research papers to 'Mr. Abc'.

3) *Hybrid Approaches*

Hybrid approaches unite both collaborative and content based systems. In several cases, combining the results of any two recommendation technique like collaborative and content-based filtering proves to be more proficient. It can be done by making collaborative and content based recommendations separately and then integration them or by accumulation of content based to collaborative approach or vice versa.

IV. WEB DATA MINING IN E-LEARNING

Web data mining is the merging of data mining and web technology. In E-learning by mining web server log records, the web data mining can discover students access patterns, such as which pages are most visited by the students, the number of student visited the page, the places from where they visited and what is the likely next page student will visit after visiting the current page[13,14,15]. The applications of web data mining in E-learning are: Personalized service, system improvement, site structural modifications, intelligent service and personalized recommendation.

A. *Personalized service*

The personalized services will be provided by attaching pages dynamically to the visited current page of the student with the other pages chosen by the other students. By mining the single students browsing information, its visiting frequency, contents length of visit, it is used to find the students interest and preferences.

B. *System improvement*

For measuring the students' contentment, the performance of web services and service quality are the key indicators. For an improved design of the web server, load balancing, data distribution, page caching and network transmission, web visiting information mining can provide valuable knowledge.

C. *Site structural modifications*

With the help of web usage mining the designers of the E-learning resources can improve the structure of the E-learning website by mining the students' feedback and their navigational path, as which page is mostly accessed by the students and how the pages are linked together. An optimized, reorganized and structured E-learning resource always attracts students.

D. *Intelligent service*

The services provided by the asynchronous E-learning platforms it should be added more intelligent factors so as to understand the students learning intent by mining their visiting behaviour. The intelligent information service can identify and predict the students' interest and hobbies according to their browsing pattern and information inquiries.

E. *Personalized recommendation*

When a student visits an E-learning resource for collecting information, they are interested in some pages of their interest but they have to visit all pages to find the desired information. During their visit the web server logs are recorded and after mining through that web server logs, interest of the student can be evaluated. The system will classify the students in different classes based on their previous visits as different students have different sets of recommendations, thus when student again visits the site a personalised page of recommendation is provided to the student.

V. THE PROPOSED FRAMEWORK

The proposed framework is based on the method proposed by F. Wang [10] that integrates data mining techniques to extract knowledge for precise E-learning applications from students historical activity, in particular for intelligent personalized services. As shown in figure. 3, the proposed framework consists of six research tasks that have to be contract with appropriately.

A. *Learning activity design*

The primary is learning activity design, which deals with the difficulty of designing learning tasks that a student is fascinated. The aim of learning action depends on the purposes of the definite E-learning applications. For instance, to assist the recommend of suitable E-learning resource, dynamic document browsing model possibly will be constructed from the browsing precedent data such that the browsing patterns of students can be reflected in an improved E-learning resource. One more instance of E-learning application is the identical of a student with some other healthy performing group of student that allocate similar activity features such that their behaviour can be referenced to give appropriate suggestions to the student. Student browsing model could be build from logged student browsing data that could replicate the real browsing behaviour of well performing

students. In outline, what learning activities are presented to students need to be resolute first before we can go on to the next step.
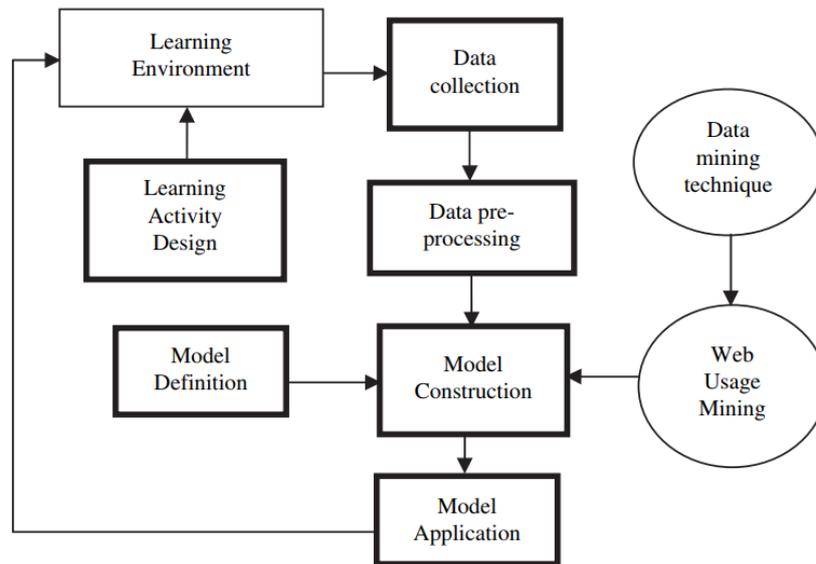

Fig.3 Proposed Architecture framework for data mining and E-learning

B. Data collection

Data following is required to make possible the data collection for the concerned learning tasks. A designer of data follower requests to select appropriate following aspects such as the students' user id, the date and time period of the learning task a student performed, and so on.

C. Data pre-processing

In this field data pre-processing is one of the important research tasks. Data has to be cleaned and transformed properly before it can be analyzed. It is necessary to convert the data to and suitable form for resolving a definite educational problem. This includes choosing what data is to bring together, focusing on the questions to be answered and making sure the data support with that questions. This is useful to retrieve this appropriate data set from raw web log records, to which different data mining and statistical techniques could be applied.

D. Model construction

Some researchers have anticipated heuristic methods to segregate a user session into a set of shorter meaningful sub-sessions. These sub-sessions are the real sessions fed into the next step of model construction using data mining techniques.

E. Model definition

The subsequently step is the model definition. Researchers have to characterize a behaviour model that imitates the real topics or concern they are engrossed. For example, a browsing model that describes the real browsing behaviour can be distinct in terms of the association and sequential browsing patterns that arise often in the history. These models can be used by teachers to discover some interesting or surprising learning patterns in students' browsing structure, and therefore might grant knowledge for teachers to restructure their content structure in a more valuable approach.

F. Model application

At the end it comes model application phase to examine the suitability and efficiency of the formulated model. Some estimation metrics have to be defined. In the following, it is presented two case studies that are based on the proposed research framework. One is for content structure model construction [16], and the other is for navigation guidance by personalized recommendation [17].

VI. CONCLUSION

Data mining in E-learning is still in its emerging phase and desires much more research efforts to make it of practically useful for teachers and students. Web mining is the sequence of task used for mining or extracting useful information from the web pages or websites. The paper has been an attempt in providing the enthusiasm toward advancing the conventional educational process by the use of data mining technology. A set of tools that

are based on data mining techniques such as clustering, and association mining, combined with collaborative filtering techniques are discussed. Then a framework is presented as a directive for online educational resources to advance their decision-making processes. It can be used to analyze the existing work, identifying existing gaps and further works. The information generated after the implementation of data mining technique is helpful for teachers as well as for students.

REFERENCES

[1] Félix Castro, Alfredo Vellido, Àngela Nebot and Francisco Mugica, "Applying Data Mining Techniques to E-Learning Problems", Applying Data Mining Techniques to E-Learning Problems, Studies in Computational Intelligence (SCI) 62, pp.183–221 (2007), Springer-Verlag Berlin Heidelberg.

[2] Sunil and Prof. M. N. Doja, "Web Usage Mining Techniques to Improve the Capabilities of E-learning Websites and Blogs", International Journal of Advanced Research in Computer Science (IJARCS), Volume 8, No. 5, May-June 2017, ISSN No. 0976-5697.

[3] Chetna Kaushal and Harpreet Singh, "Comparative Study of Recent Sequential Pattern Mining Algorithms on Web Clickstream Data", Power, Communication and Information Technology Conference (PCITC), IEEE-2015, pp.652-656, doi:10.1109/PCITC.2015.7438078.

[4] 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.

[5] P. Veeramuthu, Dr. R. Periyasamy and 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.

[6] Shaveen Singh and Sunil Pranit Lal, "Using Feature Selection and Association Rule Mining to Evaluate Digital Courseware", Eleventh International Conference on ICT and Knowledge Engineering, IEEE-2013, pp.1-7, doi:10.1109/ICTKE.2013.6756286.

[7] R. Cooley, B. Mobasher, J. Srivastava, " Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, 1(1), 1999.

[8] 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.

[9] Sunil and 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] F. Wang, "On using data mining for browsing log analysis in learning environments", WIT Transactions on State of the Art in Science and Engineering, Vol. 2,2006 WIT Press, ISSN 1755-8336 (on-line), doi:10.2495/1-84564-152-3/04.

[11] Hadeel Alharbi, Ashoka Jayawardena and Paul Kwan, "Social recommender system for predicting the needs of students/instructors: Review and Proposed Framework", International Conference on Future Internet of Things and Cloud, IEEE 2014, doi: 10.1109/FiCloud.2014.93.

[12] Harshit Patel and Prof. Pooja Jardosh, "A Survey on Recommendation System Using Web Usage Mining" International Institution for Technological Research and Development Volume 1, Issue 1, 2015.

[13] Xinjin Li, Sujing Zhang, "Application of web usage mining in e-learning platform", International Conference on E-business and E-government, IEEE 2010, doi:10.1109/ICEE.2010.353., pp. 1391-1394.

[14] 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.

[15] Sunil and Prof. M. N. Doja, "Web Data Mining in E-services–Concepts and Applications", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 8 No. 3 Jun-Jul 2017, pp.313-318, ISSN: 0976-5166.

[16] Wang, F.-H. & Shao, H.-M., Effective personalized recommendation based on time-framed navigation clustering and association mining. Expert Systems with Applications, 27(3), pp. 365–377, 2004.

[17] Wang, F.-H., On analysis and modelling of student browsing behaviour in web-based asynchronous learning environments. Lecture Notes in Computer Science, 2436, pp. 69–80, 2002.