IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.
www.ijiset.com
ISSN 2348 – 7968

# A Study on Web Usage Mining For Web Based Adaptive Educational System

**Dr. M. Durairaj[#1], C. Suresh [*2]**

[#1]Assistant professor, School of Computer Science Engineering and Applications, Bharathidasan University, Tiruchirappalli, India

[*2]Research Scholar, School of Computer Science Engineering and Applications, Bharathidasan University, Tiruchirappalli, India

**Abstract**

Education is the process of acquiring knowledge through learning. In modern era, technology based education via computer and internet gained the attention of the users. Data mining involves prominently in recent educational system to improvise the knowledge. Studies proved that data mining develops students learning behavior in a great sense. The reason behind this is its ability to evaluate and exposing the hidden data that causes success or failure. Web usage mining falls under one among the categories of data mining, is the process of extracting useful information from server logs as per the interest of the user. The objective of our paper is to review recent research papers which are exploring the significance of web usage mining in educational system. For this study, we selected recent papers from IEEE, Elsevier, ACM and other peer reviewed journals.

**Keywords:** *Data mining, Educational system, Web usage mining.*

## 1. Introduction

Education in its general sense is a form of learning in which the knowledge, skills, and habits of a group of people are transferred from one generation to the next through teaching, training, or research. Education frequently takes place under the guidance of others and also autodidactic for learners. Any experience that has a formative effect on the way one thinks, feels, or acts may be considered educational. Education is commonly categorized into stages such as preschool, primary school, secondary school and then college, university or apprenticeship.

Data mining (DM) is a computer-based information system keen to scan massive data repositories, generate information, and discover knowledge. The meaning of the traditional mining term biases the DM in grounds. But, instead of searching natural minerals, the target is knowledge. DM pursues to find out data patterns, organize information of hidden relationships, structure association rules, estimate unknown items' values to classify objects, compose clusters of homogenous objects, and unveil many kinds of findings that are not easily produced by a classic CBIS (Center for Biotechnology and Interdisciplinary Studies). Thereby, DM outcomes represent a valuable support for decisions-making [6].

With respect to education, DM focuses on knowledge discovery, decision making and recommendation. Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered that the Web usage mining is very much involved in educational arena. The term collaborates data mining and Educational system is known as Educational Data Mining (EDM). EDM suggests prototypes to inculcate, evaluate the knowledge perceived by students through various techniques

In this paper, recent research articles published on web usage mining on educational systems are reviewed which emphases the importance of DM and its role in educational research. This paper is organized as follows Section 2 contains reviews on selected papers. Section 3 describes brief explanation on each study. Section 4 contains the Mining Techniques used and Section 5 contains results and discussion. Section 6 briefs conclusion and future works.

## 2. Mining Algorithms

### 2.1 Clustering

Clustering is a division of data into groups of similar objects. Representing the data by fewer clusters achieves simplification. It models data by its clusters. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. From a machine learning perspective clusters correspond to hidden patterns, the search for clusters is unsupervised

learning, and the resulting system represents a data concept. From a practical perspective clustering plays an outstanding role in data mining applications such as scientific data exploration, information retrieval and text mining, spatial database applications, Web analysis, CRM, marketing, medical diagnostics, computational biology, and many others. Different usage of clustering technique are illustrated in [4][3][2][1].

### 2.2. Classification

Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. Classification tries to discover relationships between the attributes that would make it possible to predict the outcome. [4][2] use Classification technique.

### 2.3. Association Rules

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. [1] Uses this method.

### 2.4. Decision Trees

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. This page deals with decision trees in data mining [5].

## 3. Process in web mining usage

Web mining usage in the development of web based adaptive educational system is a laborious activity. The developer, usually the course teacher, has to choose the contents that will be shown and determine the most appropriate content elements for the types of potential user of the course. The results of these complexity decisions, taking design in one-shot is hardly feasible, even when carefully done. In many cases probably need evaluation and possibly modification of course content, structure and navigation based on students' usage information, preferably even following a continuous empirical evaluation approach. To facilitate this, data analysis methods and tools are used to observe students' behaviour and to assist instructors in detecting possible errors and shortcomings and in incorporating possible improvements. Traditional data analysis in adaptive educational system is hypothesis or assumption driven. It is very difficult for the user to find more complex patterns that relate to different aspects of the data. An alternative to this traditional data analysis is to use data mining in an inductive approach to automatically discover hidden information present in the data. Web mining, in contrast, discovers the hypothesis automatically from the extracted data rather than research-based or human-driven. It discovers interesting patterns and tendencies in student's usage information. The mined knowledge enters the system and guide, facilitate and enhance learning as a whole, not only turning data into knowledge, but also for decision making.

The web usage mining in web based adaptive educational system process consists of four steps like the process of general data mining as follows and depicted in Figure 1.

**Data collection:** The CMS system is used by students and the usage and interaction information is stored in the database. This paper discusses the usage of students' data in the Moodle system.

**Data preprocessing:** Data is cleaned and transformed into an appropriate format to be mined. In order to preprocess the Moodle data, we can use a database administrator tool or some specific preprocessing tool.

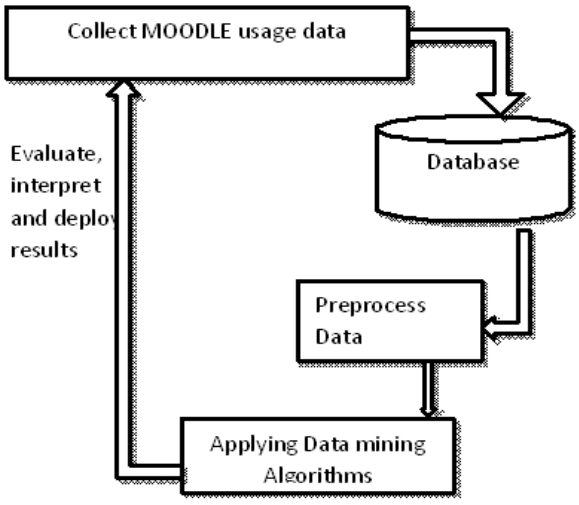


**Fig. 1. Web usage mining in Moodle data**

ISSN 2348 – 7968

## 4. Web mining usage in adaptive educational system

Recent research articles on web usage mining for web based adaptive educational system were taken from a leading full text scientific database called ScienceDirect.com. The journals were published during the years 2008 to 2014. Each paper specifies the impact that they makes towards educational data mining researchers and the field itself. Let us begin with a brief representation of each paper in a table below.

| Ref. No. | Authors | Objective | Platform | Data Mining Task | Number of Citations |
|---|---|---|---|---|---|
| [4] | Cristóbal Romero, et. al. | Describes the process of mining e-learning data with the techniques of Moodle data. | Moodle - an Open Source Course Management System | Clustering and Classification | 80 |
| [2] | Cristóbal Romero, et. al. | Proposes the usage of different data mining approaches for improving prediction of students' final performance based on participation indicators in quantitative, qualitative and social network forums. | On-line discussion Forum | Classification and Clustering | 80 |
| [3] | Cristóbal Romero, et. al. | Proposes an advanced architecture for a personalization system to facilitate Web mining. | AHA - a well-known open source general-purpose adaptive hypermedia system. | Clustering | 28 |
| [1] | Feng-Hsu Wanga, et. al. | Proposes a new clustering method called HBM (Hierarchical Bisecting Medoids Algorithm) to cluster users based on the time-framed navigation sessions | E-learning | Clustering and Association rule Mining | 19 |
| [5] | Srecko Natek, et. al. | Focuses on predicting the success rate of the students in Higher Education Institutions | MS Excel Tool | Decision Trees | 26 |

## 5. Review of chosen literatures

Cristóbal Romero, et. al. [4] developed a Course Management System, which stores and uses the data collected the student's usage and interaction data from Moodle system. The collected data was cleaned and transformed into an appropriate format to be mined. The data mining algorithms were then applied to build and execute the model that discovered and summarized the knowledge of interest to the user (instructor, student and administrator). The results or model obtained were interpreted and could be used by the instructor for further actions. The instructor could use the information discovered to make decisions about the students' and Moodle course activities to improve the students' learning tool.

Cristóbal Romero, et. al. [2] for experimenting Moodle platform, data was gathered from 114 university students during a first-year course in computer science in 2011–2012. This was an introduction course to computer science from a theoretical and practical point of view that is entitled "Computer Science Fundamentals". The course was enhanced by using a Moodle platform for providing the students with supplementary on-line resources, activities, and a discussion forum. The instructions given to the students were to use the forum for discussing course contents, solving doubts and problems between students in which the instructor would not participate in the discussion. Students could ask questions about theory or exercise problems, replies to previous messages or just to browse through the discussion. The idea was that some students could help other students by using the discussion forum. Although it was not mandatory to join the forum, in order to encourage students to use the forum, the instructor also remarked that the level of participation in the forum would have a positive effect in the case of students that obtain a near pass mark in the final exam. Finally, at the end of the course, students carried out a final pen and paper exam to evaluate them, and of the114 students: 68 passed (59.65%) and 46 failed (40.35%) the course. The data mining approach applied for improving the prediction of students' final performance from

forum data used in this system is as shown in Figure 2.

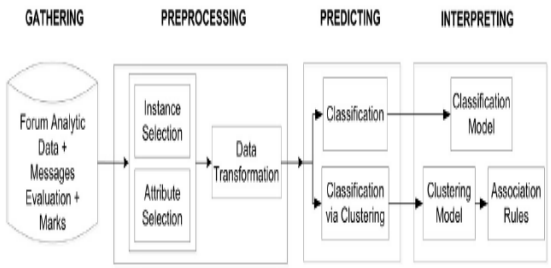


**Fig. 2. Data mining approach [2]**

Cristóbal Romero, et. al. [3] has proposed architecture of a recommender system that utilizes Web usage mining to recommend the links to visit next in an adaptive Web-based educational system. A specific mining tool and a recommender engine have been developed to help the instructor to carry out the Web mining process. Although they have integrated both the Web mining tool and the recommender engine into the AHA! System, it can, in principle, also be used in other Web-based educational systems. AHA system has some adaptive hypermedia methods and techniques that are especially useful for educational applications, which are User model based on concepts and Adaptive link hiding or link annotation. Author tested AHA architecture and algorithms proposed with several experiments. The data used in this study were collected from the online AHA! tutorial (http://aha.win.tue.nl/tutorial/) that consists of 43 Web pages. For experiments, author used a total number of 78 students with 118 sessions and 684 records. These students are mainly TU/e (Eindhoven University of Technology) students taking a traditional course in adaptive hypermedia and some other Internet users interested in the AHA! system and taking the tutorial online.

Feng-Hsu Wanga, et. al. [1] proposed a personalized recommendation scheme. The recommendation module builds a knowledge base of navigation patterns by first clustering users based on the time-framed navigation sessions over the historical navigation database, and then establishes the access patterns for each user group using the association-mining technique. To produce personalized recommendations for a user, the group most similar to the user's navigation sessions is first selected, and then the recommender applies the prediction rules in the corresponding rule base to generate the item recommendation list that sorts the items in terms of relevance. This work concluded that the proposed architecture can discover as many rules as a basic architecture but with higher values of confidence and support. Historical navigation data was collected from three classes (classes A, B and C) of a virtual classroom course ('Expert System') for one semester. Data was preprocessed for experiments. Three sets of experiments with different amounts of training data were conducted. The first one used a half of the session data from class A for training, and the other half for testing. The second one used the whole data of class A for training, and the whole data of class B for testing. The last experiment used the data of class A and B for training, and used the data of class C for testing. The obtained results were compared.

Srecko Natek, et. al. [5] developed a system in which the first step was data requirements identification where HEIs require predicting the student's success rate related to the specific course. The data was imported from the web application. Validate, explore, clean and transpose data were the next activities where irrelevant columns were deleted (e.g. group, remark, individual activities points, registration number) and columns order were adjusted accordingly (logical order for success prediction, the last column was set as the predictable attribute: ''Final grade''). In the next technological step the Data Mining Technology were chosen. After this step where the Data Mining modeling Technique were chosen, finally the best model was chosen and the results were evaluated. The study explores the possibility to predict the success rate of students enrolled to an academic course using a contemporary data mining tools normally available to HEIs. The research clearly exhibits that available desktop data mining tools have matured in terms of their usability and ease of use, and provide usable results without extensive investment.

Olga C. Santosa, et. al.[7] address an open key issue during the development of web-based educational systems. In particular, author provided an educational-oriented approach for building personalized e-learning environments that focuses on putting the learners' needs in the center of the development process. Their approach proposes user centered design methodologies involving interdisciplinary teams of software developers and domain experts. It was illustrated in an adaptive e-learning system, where a MOOC (Massive Open Online Course) was taken by nearly 400 learners. In particular, authors' report where user centered design methods could be applied along the e-learning life cycle to designing and evaluating personalization support through recommendations in learning management systems.

Brusilovsky, P., & Peylo, C [9] demonstrated an adaptive and intelligent web-based educational system, which provide an alternative to the

traditional just put-it-on-the-web approach in the development of web based educational courseware. This was attempt to be more adaptive by building a model of the goals, preferences and knowledge of each individual student and using this model throughout the interaction with the student in order to adapt to the needs of that student. This system was result of a joint evolution of intelligent tutoring intelligent tutoring systems (ITS) and adaptive hypermedia systems (AHS). Some examples of ITS are SQL-Tutor, German Tutor, ActiveMath, VC-Prolog-Tutor, and some examples of AHS are AHA!, InterBook, KBS-Hyperbook, WebCOBALT [9]. The data from this system are semantically richer and can lead to more diagnostic analysis than data from traditional web-based education system. Data collected from the domain model, which may be structured into ontology, pedagogical dataset, interaction log files and student model.

Chiao-Tzu Huang a, et. al. [10] conducted a conducted a random sampling of 38,000 pieces of data in the personnel educational training database of China Motor Corporation according to employees' positions and departments for innovation and breakthrough. This was done through the decision tree algorithm for study. Mining and analysis were done in order to find the relation between the classification of educational training courses and the classification of the employees' occupation, the courses favored by the employees of different occupations and their accomplishment probability.

## 6. Discussion and Comparative Analysis

In [2], predicting final marks were used in predicting students' success or failure in a course. This is a classification problem and not a regression problem and applied DM methods. On the other hand, they propose to use classification via clustering as an alternative method. Clustering and classification are both classification methods, although clustering is an unsupervised method and classification is a supervised method. Clustering algorithm was firstly executed using the training data, after removal of the class attribute, and the mapping between classes and clusters was determined. This mapping was then used to predict class labels for unseen instances in the test data.

In [1], the work was to propose an effective personalized recommendation system based on time-framed navigation clustering and association mining. The experimentation results showed that the clustering recommendation methods based on the 'week' frame size could characterize the users' behavior accurately. Best average weighed precision rate is 0.6, average weighted recall rate is 0.7 and average service rate is 0.5, respectively. It shows that precision and recall rates are accurate than the conventional non-clustering one. The paper concludes that the recommendation method uses a shorter frame size such as a week for clustering user navigations and mining association rules, because a shorter frame size could track more flexibly the changes of users' traversal behavior.

Data mining algorithms and techniques have been used in educational system to mine the tasks such as sequence pattern, statistics and sequence pattern, prediction, association and classification, visualization, text mining, outlier detection, etc. [8]. Data mining was used to know how students use the course, how a pedagogical strategy impacts different types of students, in which order the students study subtopics, what are the pages/topics that students skip, how much time the students spend with a single page, a chapter or the full course, etc.

In [9], adaptive and intelligent web-based educational systems attempted to be more adaptive by building a model of the goals, preferences and knowledge of each individual student and using this model throughout the interaction with the student in order to adapt to the needs of that student. This system used a standard student model, internally by the tutoring system, but, for the purpose of data mining, it developed a new model of student interaction with augmented information with contextual data.

In [10], the analytical framework in this work was based on Six Sigma management and the action steps of Six Sigma when introducing the projects. Artificial neural network was used and it was optimized with different standards. Levenberg Marquardt back-propagation (LMBP) was used for function approximation and it reveals considerably precise training. Compared with other algorithms, LMBP acquired lower mean square error. In the input layer, there are 6 variables such as educational level, age, work year, position, department and gender are used and these could influence training classification. This work concluded as they could find that there were more production and R&D courses in the China Motor Corporation and fewer marketing, human resources and finances. The work suggested that the company could increase the number of courses in the latter category.

## 7. Conclusion and Future work

The study answers the research questions and supports the conclusions that web usage mining in Adaptive educational system has a real potential and need of the hour. Currently, most of the

researches are being carried out on educational data mining which entices attentions towards the use of e-learning tools such as Moodle, WebCT, AHA, MOOC and other customized tool for the learning purposes. Discussed study reveals that the quality of the educational process is significant in educational institutions. This study clearly supports that the web usage mining is not only limited to large data sets as the majority but structured small data set can also serve the usable results.

Data mining for small student data sets from research is relevant example of effective use of web usage mining technology to develop Adaptive educational systems in the education domain. In the past, there were several attempts to predict the student educational success with various successes. Some of the researchers focused on small student data sets and the results obtained were appeared to be promising.

We hope that this review will be able to shed some useful insights for researchers and educators in order for educational data mining to become a mature area. This study advocate small data sets web usage mining possibilities on the theoretical level. As we considered, the research can be further expanded to develop an e-learning website for our university computer science department. We hope that these conclusions will encourage web usage mining applications to develop an Effective Adaptive Educational System.

## 8. References

[1] Feng-Hsu Wanga, Hsiu-Mei Shaob, "Effective personalized recommendation based on time-framed navigation clustering and association mining", Expert Systems with Applications, Vol. 27, Pp. 365–377, 2004.

[2] Cristóbal Romero, Manuel-Ignacio López, Jose-María Luna, Sebastián Ventura, "Predicting students' final performance from participation in on-line discussion forums", Computers & Education, Vol. 68, Pp. 458–472, 2013.

[3] Cristóbal Romero , Sebastián Ventura , Amelia Zafra , Paul de Bra, "Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems", Computers & Education, Vol. 53, Pp. 828–840, 2009.

[4] Cristobal Romero, Sebastian Ventura, Enrique Garci, "Data mining in course management systems: Moodle case study and tutorial", Computers & Education, Vol. 51, Pp. 368–384, 2008.

[5] Srecko Natek, Moti Zwilling, "Student data mining solution–knowledge management system related to higher education institutions", Expert Systems with Applications, Vol. 41, Pp. 6400–6407, 2014.

[6] Alejandro Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works", Expert Systems with Applications, Vol. 41, Pp. 1432–1462, 2014.

[7] Olga C. Santosa, Jesus G. Boticarioa, Diana Pérez-Marín, "Extending web-based educational systems with personalized support through User Centred Designed recommendations along the e-learning life cycle", Science of Computer Programming, Vol. 88, Pp. 92–109, 2014.

[8] C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications" Vol. 33, Pp. 135–146, 2007.

[9] Brusilovsky, P., & Peylo, C., "Adaptive and intelligent web-based educational systems", International Journal of Artificial Intelligence in Education, Vol. 13, Pp. 156–169, 2003.

[10] Chiao-Tzu Huang a, Wen-Tsann Lin a, Shen-Tsu Wang b, Wen-Shan Wang a, "Planning of educational training courses by data mining: Using China Motor Corporation as an example", Expert Systems with Applications, Vol. 36, Pp. 7199–7209, 2009.