A Survey of Approaches and Tools Used in Educational Data Mining

D. Gandhimathi[1] S. Gomathi[2]

Assistant Professor, Dept. of Computer Science and Application and Software Systems, Sri Krishna Arts and Science College, Kuniamuthur, Coimbatore, Tamil Nadu, India[1]

M.Phil. Scholar, Dept. of Computer Science and Application and Software Systems, Sri Krishna Arts and Science College, Kuniamuthur, Coimbatore, Tamil Nadu, India[2]

ABSTRACT: Educational Data Mining (EDM) is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn. EDM uses a computational method to examine the educational data and is concerned with developing methods to explore the unique types of data in educational fields. For this data exploration, many tools were used like personal learning environments and recommender systems, Context learning and Course management systems. These tools provide various benefits for the educational data mining. For this exploration of the data, in this survey we focus and provide a various tools of research trends using EDM Tools to explore data.

KEYWORDS*:* EDM, EDM tools, Exploring data, Knowledge discovery and Learning analysis.

I. INTRODUCTION

Present advanced computer technologies like machine learning and data mining systems are using different data extraction methods because extracting educational data is significant one, where data set are broadly used. Educational data mining (EDM) is a field that utilizes statistical, machine-learning, and data-mining (DM) algorithms over the different types of educational data. Its main purpose is to consider these types of data in order to determine educational research issues. EDM is troubled with developing methods to discover the unique types of data in educational environment and, using these methods, to improved understand students and the scenery in which they be trained. On one hand, the enlarging in both involved educational software as well as state databases of student's information have shaped large repositories of data shimmering how students learn. On the other hand, the use of Internet in education has created a new context recognized as e-learning or web-based education in which huge amounts of information about teaching–learning interface are continuously generated and universally available [1]. DM methods, admired in the field of EDM, can be divided into the following common categories: prediction, clustering, relationship mining, discovery with models, and data retrieval for human judgment. The first three types are measured to be universal techniques across all types of DM tasks. The fourth and fifth categories are useful in the field of EDM. The EDM provides novel possibilities for gathering, analyzing, and presenting data. These new data sources can be utilized as guides for course redesign or as an evidence for implementing new assessment approaches. Researchers at the University of Auckland developed a modeling tool facilitating to suggest novel assessments through network maps of online discussions. The modeling tool becomes much more efficient to see how new threads emerge with snapshots of the discussion. It also makes it easier to pinpoint discussion leaders. The graphical representations of the timestamp and of the posts network help teachers to assess students according to the interactions in discussion groups. The information gained from LMS can be directly used as a significant indicator of the potential student failure. It enables to design an alert system [2]. The actions can be performed with data reporting or from a predictive model. Students can change their study habits when the early

student assessment data is available. It can affect student motivation or even retention. It is a commonly used method for motivating of under-performing students. They need to be oriented to the relevant help.
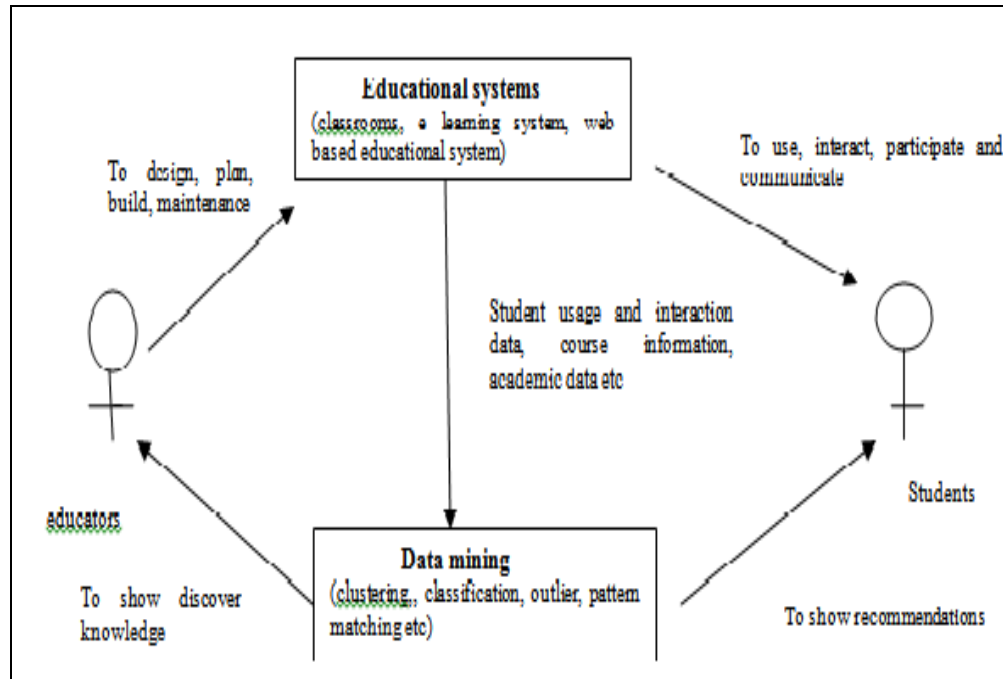

Fig 1: Education System in Data Mining

Educational data mining means discover hidden data that invent from educational settings by using new methods for better explanation of students and settings they learnt. Educational data mining promote distinct tools and algorithms for analyze the data patterns [3]. In EDM, data is accumulated during learning process and then study can be done with the techniques from statistics, machine learning and additional data mining concepts. To extract the hidden knowledge from data came from educational system, the different data mining techniques like classification, clustering, rule mining etc. There is pressure in higher educational institutions to provide up-to-date information on institutional effectiveness. Institutions are also increasingly held accountable for student success. One response to this pressure is finding new ways to apply analytical and data mining methods to educationally related data [4]. This Survey shows the various tools used for the education data mining to provide a different and best result for the education field [5].

II. LITERATURE REVIEW

Even though data mining (DM) has been functional in numerous industry and sectors, the submission of DM to educational contexts is limited. Researchers have establish that they can relate data mining to rich educational data sets that approach from course management scheme such as Blackboard, Angel, WebCT, and Moodle. The rising field of educational data mining (EDM) examines the exclusive ways of applying data mining methods to answer educationally linked problems. The aim of this survey is to afford a comprehensive study of various researchers' approaches and their limitations for this exploring education data mining.

Hijazi and Naqvi [6] conducted as study on the student performance by selecting a model of 300 students (225 males, 75 females) from a group of colleges associated to Punjab university of Pakistan. The hypothesis that was stated as "Student's approach towards attending in class, hours depleted in study on daily basis after college, students' family earnings, students' mother's age and mother's education are appreciably linked with student performance" was casing. By means of simple linear regression analysis, it was found that the factors like mothers education and students family income were highly associated with the student university performance.

In this paper [7], we gave a case study in educational data mining. It illustrated how useful data mining can be in higher education in predominantly to progress student routine. We used students' data from database course. We collected all accessible data together with their usage of Moodle e-learning facility. It functionally data mining techniques to discover the knowledge's. Predominantly we discovered association rules and we arrange the rules using lift metric then we envisage the rules. Then we exposed classification rules using decision tree. Also we clustered the student into group using EM clustering. Finally, using outlier examination we detected all outliers in the data. Each one of this acquaintance can be used to recover the performance of student. Also, experiments could be done using more data mining techniques such as neural nets, genetic algorithms, k-nearest Neighbor, Naive Bayes, support vector machines and others. Finally, the used preprocess and data mining algorithms could be surrounded into e-learning system so that anyone using the scheme can benefit from the data mining techniques.

This paper surveys [8] how the computer input/output (I/O) subsystem is taught in opening undergraduate courses. It is significant to study the educational development of the computer I/O subsystem because, in the curriculum recommendations, it is measured a core topic in the area of knowledge of computer architecture and organization (CAO). It is also a basic knowledge to be attained in order to work in areas such as human–computer interaction (HCI) or embedded systems. Examination questions, course syllabi, and textbooks were analyzed to make out which teaching advance is being used. Individuals teaching the I/O subsystem could choose between the options make clear here, according to their proposed learning outcomes. In addition, a literature survey was demeanor on the development and use of tools to recover student considerate of I/O and to make the subject less abstract and more nice-looking [9]. A goal is to point toward to computing education researchers that the preponderance of the literature reports understanding in mounting or using different resources or educational methodologies, but that these are not based on a conjecture of learning.

This paper [10] reports accomplish efficient, generic explanation to a major rising problem in educational data mining: competent exploration of enormous student-tutor contact logs. We spot several useful necessities for a tool to support such searching. Our key conceptual donation uses temporal relations to representation natural hierarchical structure. This is the intellect in which "time will tell" many basic associations among tutorial events. Structural confirmation of generality includes the tool's mostly tutor-independent intend, imitate in the code's conciseness and its insufficiency of references to detailed tables or fields of the database. Empirical evidence of sweeping statement includes triumphant use of the tool with databases from different years' versions of the Reading Tutor. It is early to appraise usability because the tool is still so new. However, we can claim a ten- or hundred-fold lessening in keystrokes compared to obtaining the same information by question the database directly [11]. For example by clicking on an item in the event list display its context as a chain of ancestor events. Recognize these ancestors by query the database directly would necessitate querying a separate table for each ancestor.

Matthew W. Johnson, et al [12] presented EDM Vis a software hallucination tool calculated for exploring and interacting with software instructor log data. Our tool allows users to get a better understanding, overlooking strictly at tutor logs, of how students work together with a software tutor by visualizing the tutor's interaction logs. EDM Vis is our first steps to provided that a domain independent visualization tool for understanding student performance in software tutors and our original results seem shows potential for the future expansion of this tool. consent to the user to annotate and supplement their data in the visualization seems useful, predominantly for highlighting an important characteristic of the data a user wishes to share [13] In addition a method of allowing assessment between problems could prove appealing, for both

looking at a single student over multiple troubles as well as the same problem from multiple tutors. It is also understandable that for each graph and sub-graph we should make available some general statistics to deliver the user with existing numbers, fundamentally more details on demand. In calculation, we must fix some problem with our use of color, making it easier to know which nodes are errors when select. These types of collection and filtering and their addition seem to be a good course to pursue for the next iteration of EDM Vis. Lastly; the counsel we gathered from the qualitative segment of the survey will definitely be taken into account and put into practice in all possible cases.

III. VARIOUS APPROACHES USED IN EDM

In EDM, data is mounting up during learning procedure and then study can be done with the practice from statistics, machine learning and other data mining impression To extort the hidden knowledge from data came from educational system, the different data mining techniques like classification, clustering, rule mining etc. (EDM) examines the exclusive ways of applying data mining methods to solve educationally associated problems. Here we show the diverse research approaches for the education data mining.

a. Descriptive Approach: In this approach, only an explanation of the computer I/O subsystem is given to the students, who are then probable to be able to describe the concepts. At most, they could be asked to identify associations among concepts. This is an easy way to initiate the topic that could even be used with students who are not majoring in calculate

b. Offline Education: It tries to convey knowledge and skills based on face-to-face contact and also study psychologically on how humans learn. Psychometrics and statistical method have been practical to data, like student's behavior/ performance, curriculum, etc., that was gathered in classroom environments.

c. E-learning and learning management system (LMS): E-learning makes available online instruction, and LMS also provides communication, collaboration, administration, and reporting tools. Web mining (WM) practice has been practical to student's data accumulate by these systems in log files and databases.

d. ITS and AEHS: Intelligent tutoring system (ITS) and adaptive educational hypermedia system (AEHS) are a substitute to the just put-it-on-the-web approach, trying to become accustomed teaching to the needs of each meticulous student. DM has been applied on the data those selected by the systems such as log files, user models, etc.

IV. VARIOUS TOOLS USED IN EDM

a. Personal Learning Environments

Personal Learning Environments (PLE) and Recommender Systems (PRS) are the related concept to the educational data mining. PLE provide a various tools and services to the EDM that adapts the students to learning in personal. Nowadays Students want to learn their subject related concepts through personally without interaction of the teacher's students learn their subjects through the internet. PLS is a tool that prefers the online learning technique which provide online tutorial. Student can gather the data or information from online. It helps the students to interact with their subjects easily [8].

b. Personal Recommender Systems

Personal Recommender Systems (PRS), must be adapted when they are used in educational contexts because the recommendations should coincide with enlightening objectives. The cause is that it is not possible to apply existing recommender systems honestly to educational data because they are highly domain dependent. There are two significant challenges with respect to applying recommender systems in an educational context. At first, the system must challenge to understand or conclude the needs of learners. Second, there must be some way for ability members to organize

recommendations for their learners. This system support learners by providing them with highly individualized recommendations for improved learning efficiency.

c. DLSim3: It is GUI-based digital logic simulator (DLSim3) with a collection of circuits and plugins (DLsys) casing all aspects of the computer architecture. The display place can be found in [31] and contains several plugins that reproduce I/O devices. They also suggest some exercises. As for the I/O subsystem, they suggest connecting several I/O devices to a single interrupt request line using a daisy chain or adjust the CPU plugin so as to disconnect the I/O address space from memory address space. Since the tool offers the opportunity of changing the system itself, it can be well thought-out as a representative of the data-path signal approach.

d. EDM IN COURSE MANAGEMENT SYSTEMS [8]

A large number of researchers from EDM focus on course management systems and how they can be improved to support student erudition outcomes and student success. One research team developed a simplified data mining toolkit that activate within the course management system and allows non-expert users to get data mining information for their courses. In addition, a toolkit allows teachers to work in partnership with each other and share results. This research is important because most data mining tools are complicated and require deep expertise in data mining tools, methods and processes, statistics, and machine learning algorithms. The data mining process recurrently follows a pre-processing phase, then an application of specific data mining techniques, and then a post-processing phase. Course management systems such as open source Moodle can be mined for usage data to find interesting patterns and trends in student online performance. An organized method for applying data mining techniques to Moodle usage data was established. The assistance to mining usage data is that it contains data about every user activity, such as testing, quizzes, reading, and discussion posts.

V. CONCLUSION

We analyzed several EDM tools and its features in this paper as a survey. It has been introduced as an upcoming research area related to several well-established areas of research, including e-learning, AH, ITSs, WM, DM, etc. Moreover, several tools found in literature, which have been selected based on their suitability for teaching in the field of EDM. Without obvious metrics for evaluating how different approaches/tools affect the knowledge gained by students, it is hard to rank these methods according to a clear outcome. This paper provides a good overview of the outcomes intended with each approach and tool.

REFERENCES

[1]Cristˊobal Romero, "Educational Data Mining: A Review of the State of the Art" VOL. 40, NO. 6
[2] Jan Géryk, "Applied Visual Analytics for Higher Education Decision Making".
[3] Nitya Upadhyay and Vinodini Katiyar, "A Survey on the Classification Techniques In Educational Data Mining".
 [4] Richard A. Huebner, Norwich University, "A survey of educational data-mining research", Research in Higher Education Journal.
[5] Ryan S.J.d. Baker, "Data Mining for Education", Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
[6] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting studentˮs performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
[7] Alaa el-halees, "Mining students data to analyze learning behavior: A case study".
[8] Edurne Larraza-Mendiluze and Nestor Garay-Vitoria, "Approaches and Tools Used to Teach the Computer Input/Output Subsystem: A Survey"
[9] A.Dinesh Kumar ,Dr.V.Radhika, "A Survey on Predicting Student Performance".
[10] Jack Mostow, Joseph Beck, " An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell!".
[11] Edurne Larraza-Mendiluze, "The Computer Input/Output Subsystem Education in an undergraduate introductory course: a Multiperspective Study"
[12] Matthew W. Johnson, et al, "The EDM Vis Tool" University of North Carolina at Charlotte.
[13] Marie Bienkowski, Mingyu Feng, Barbara, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief".
[14] G. Pankaj Jain, Varadraj P. Gurupur, Jennifer L. Schroeder, and Eileen D. Faulkenberry, "Artificial Intelligence-Based Student Learning Evaluation: A Concept Map-Based Approach for Analyzing a Student's Understanding of a Topic".

[15] Crist´obal Romero, "Educational Data Mining: A Review of the State of the Art"

BIOGRAPHY

D.Gandhimathi

She Completed M.Phil - Computer Science in Bharathiar university. She is specialized in the area of Database System, Datamining .Her academic experience is 9 years till date.She had presented 8 papers in national & International Conference and published 6 papers in various Journals.

S. Gomathi

She received the Master's degree in the Department of Computer Science from Bharathiar University in 2012. She has industrial experience and worked in Data ware house project and one and half years of academic experience. Her Research Interest include Data mining, Dataware House etc.