EDUCATIONAL DATA MINING APPLICATIONS

S. Lakshmi Prabha

Department of Computer Science,Seethalakshmi Ramaswami College ,Tiruchirappalli, Tamilnadu, India

Dr.A.R.Mohamed Shanavas

Department of Computer Science,Jamal Mohamed College,Tiruchirappalli,Tamilnadu, India

ABSTRACT

The tools and techniques of data mining are being adopted by the industries to generate business intelligence for improving decision making. Education institutions are beginning to use data mining techniques for improving the services they provide and for increasing student grades and retention. This paper presents broad areas of applications in which educational data mining can be applied to e-learning. The application areas discussed in this paper are:

1) User modelling
2) User grouping or profiling
3) Domain modelling and
4) Trend analysis.

A MathsTutor for school students of 6^{th}, 7th and 8^{th} grade is designed and implemented in 3 schools in Tamilnadu. The student data are taken for analysis for the above said areas. This helps teachers, policy makers, and administrators to understand how educational data mining work to support education related decision making.

KEYWORDS

Educational Data Mining, E-Learning, Profiling, Trend analysis, User and domain modeling.

1. INTRODUCTION

Educational data mining (EDM) is a research area which utilizes data mining techniques and research approaches for understanding how students learn. Interactive e-learning methods and tools have opened up opportunities to collect and scrutinize student data, to ascertain patterns and trends in those data, and to formulate new discoveries and test assumptions about how students learn. Technology enhanced learning relies heavily on learning management systems (LMS) or course management systems (CMS). These LMS/CMS automatically record the key strokes of individual users as server logs. Mining these logs provide patterns to teachers to identify slow learners and can adjust teaching strategies.

Data collected from learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms and techniques can explore for model building.

Educational data mining researchers [1][2] view the following as the goals for their research:

1. Predicting students' future learning behavior by creating student models that incorporate such detailed information as students' knowledge, meta-cognition, motivation, and attitudes.

2. Discovering or improving domain models that characterize the content to be learned and optimal instructional sequences.

3. Studying the effects of different kinds of pedagogical support that can be provided by learning software; and

4. Advancing *s*cientific knowledge about learning and learners through building computational models that incorporate models of the student, the software's pedagogy and the domain.

To accomplish these goals, educational data mining research uses technical methods like prediction, clustering, relationship mining, modeling etc. As described in a practice guide of the Department of Education's Institute of Education Sciences [3], working from student data can help educators both track academic progress and understand which instructional practices are effective. The guide describes also how students can examine their own assessment data to identify their strengths and weaknesses and set learning goals for themselves.

This paper considers the MathsTutor [6] as an e-learning environment implemented for Tamilnadu State Board syllabus for the school students of grade 6^{th} to 8^{th} in the area of mensuration. The students learn the problem through examples, understand through doing exercises and test his knowledge through test. Each of these category has three types of problems (mental, simple, big) according to number of steps required for solving and knowledge requirement. The number of problems a student solves is 60(example-20+exercise-20+test-20). Sixty students of 6^{th} grade participated in this project. So total of 3600 entries are stored along with the present status of student (like completed the learning material/completed how many problems in each category etc.) keystrokes, steps, results, feedback, number of hints used, time spent in solving a problem. This makes the data big and EDM tools are required for analysis. EDM is broadly explained in [7] to [16].

The following section describes how EDM can be applied in e-learning to improve the performance of students and institutions.

2.APPLICATION AREAS FOR EDM

Romero & Ventura[15] reviewed the EDM articles and suggested that future EDM research focus on the following aspects:

- integrate EDM tools with e-learning systems
- standardize data and models
- make EDM tools easier for educators and non-expert users
- customize traditional mining algorithms for educational context

This section briefs about the application areas for EDM in e-learning systems. The screen shots are taken from the MathsTutor prepared for Tamilnadu State Board syllabus of 6^{th} to 8^{th} grade maths.

User Modeling

User modeling encompasses what a learner knows, what the user experience is like, what a learner's behavior and motivation are, and how satisfied users are with online learning. User models are used to customize and adapt the system behaviors to users specific needs so that the systems 'say' the 'right' thing at the 'right' time in the 'right 'way [5]. EDM can be applied in modeling user knowledge, user behavior and user experience.

User Knowledge Modeling

Inferring what content does a student know like specific skills and concepts or procedural knowledge and higher order thinking skills is known as user knowledge modeling. Knowledge can be inferred from accumulated data that represent the interactions between students and the learning systems such as correctness of student responses alone or in a series, time spent on practice, number and nature of hints requested, repetitions of wrong answers, and errors made. Such "inferences" can be made by a predictive computer model or by a teacher looking at student data on a dashboard. The Figure 1 shows the
multistep solution performed by a student. The student log stores all the above said interactions between student and MathsTutor.

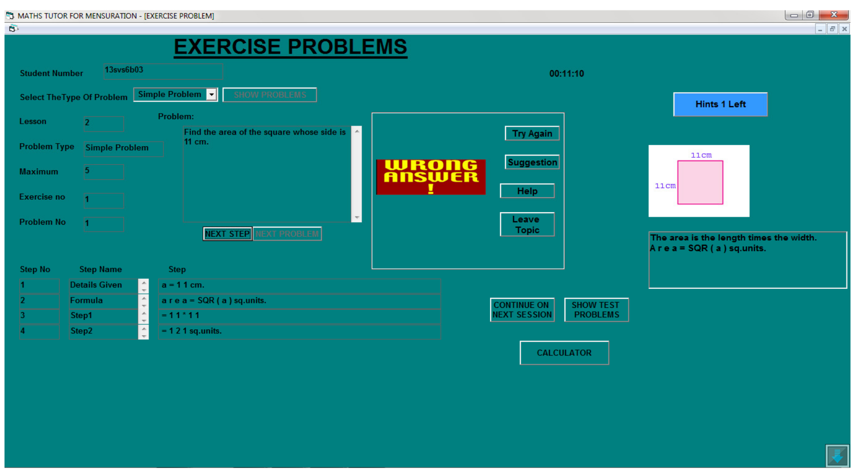

Figure 1. Screen shot from MathsTutor : Student solving exercise problem.

This modeling has been adapted to build adaptive hypermedia, recommenders systems, expert systems and intelligent tutoring systems. In intelligent tutoring systems, user knowledge models direct key operations, such as deciding which problems to give students. A popular method for estimating students' knowledge is Corbett and Anderson's knowledge tracing model [4], an approach that uses a Bayesian-network-based model for estimating the probability that a student knows a skill based on observations of him or her attempting to perform the skill.

User Behavior Modeling

Online learning systems log student data that can be mined to detect student behaviors that correlate with learning. User behavior modeling in education often characterizes student engagement. It relies on the same kinds of learning data used in predicting user knowledge plus other measures, such as how much times a student spent online (or on the system), whether a student has completed a course, attendance, tardiness, and sometimes a student's level of knowledge as inferred from his or her work with the learning systems or from other such data sources as standardized test scores.

In this MathsTutor environment it is proposed to preprocess the student log from solving exercise problem to derive new attributes from hints used, correct/incorrect step(s), suggestions received

User Profiling

Profiling means grouping related users into categories using prominent characteristics. Then, these categories can be used to provide experiences to group of users or to make recommendations to the users. In education, data mining techniques, such as clustering and classification, are often used to sort out (or profile) students based on the kinds of personal learning data, on student demographic data or both. The Figure 2 shows the classification of 60 students of grade 6 according to class label remark using J48 classifier in WEKA.

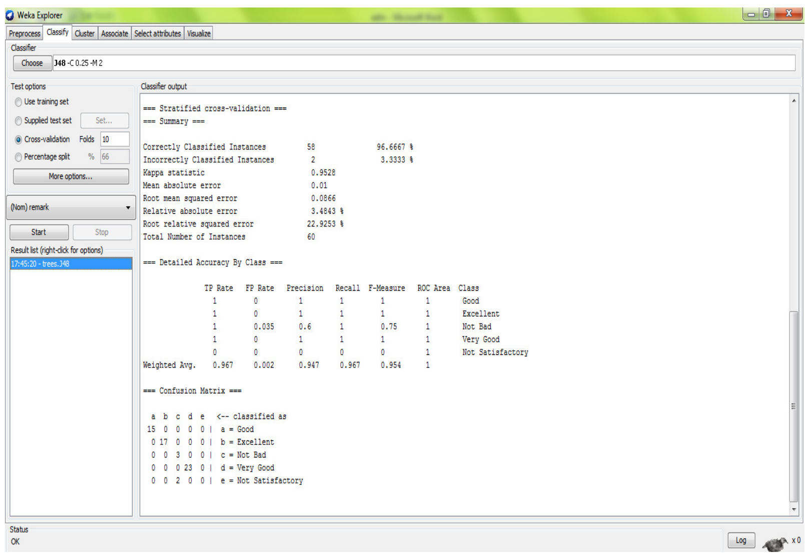

Figure 2. J48 Classifier on student Test data in WEKA.

The confusion matrix provides solution to the classification problem as 15 students got the remark Good, 17 students got the remark Excellent, 3 students got the remark Not Bad, 23 students got the remark Very Good& 2 students got the remark Not Satisfactory.

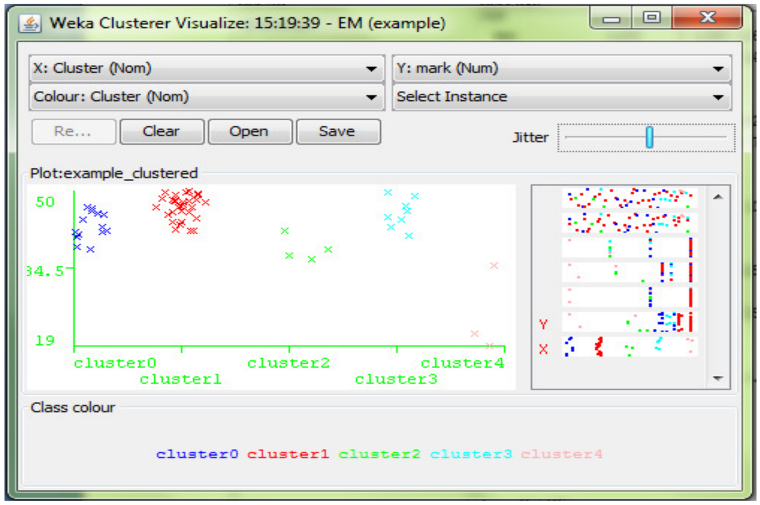

Figure 3.EM cluster visualize

The EM algorithm is used to create the cluster with a minimum size & density. This algorithm also handles the outlier problem. The result of training set using this algorithm is given in figure 3.

Domain modeling

Domain modeling is largely experimental with the goal of understanding how to present a topic and at what level of detail. A domain model is often created to represent the key concepts that make up a subject or topic area like mathematics. The domain model also identifies the relationships among all key concepts or units of study. Research in domain modeling in educational data mining investigates how learning is affected by differences in how a topic is divided into keys concepts at a particular level of generalization.

In MathsTutor, mensuration part of mathematics is taken for the study. It covers a) metric measures, b) area, perimeter and volume of solid figures (square, circle, triangle, rectangle, parallelogram, etc.), various geometrical shapes (cone, cylinder, sphere etc.)

The student module of this project covers learning material for each lesson and problems to be solved. The problems are categorized as example, exercise and test. In each category the problems are classified into Mental, Simple, and Big problems. After learning the lesson material the student has to go through example problems solved by the tutor with explanations. Then he enters into the exercise area where he has to solve the problem step by step and can get the help form tutor and receiving feedback for wrong step (Fig. 1). His understanding is checked in the test process, where he solves the problem without any help from the tutor.
The domain modeling of this project is enhanced from the feedback got from the teachers.

Trend Analysis

In education, trend analysis helps answer such questions as what changes have occurred in student learning over time and how learning has changed. At the school level, trend analysis can be used to examine test scores and other student indicators over time to help administrators determine the impacts of policies. In educational data mining, trend analysis often refers to techniques for extracting an underlying pattern, which might be partly or nearly completely hidden by data that does not contribute to the pattern (i.e., noise). This part is under study for our MathsTutor log data.

This section has described broad categories of applications that exploit educational data mining to adapt and personalize learning and improve teaching.

3.CONCLUSION

This paper discussed about the applications of EDM that made predictions and recommended actions based on increased visibility into student actions. The experiment is done on 6^{th} grade Student log collected from MathsTutor for mensuration. This paper provides only limited number of screen shots applied on student data. By identifying the knowledge level of a student and grouping them will make easier for the teacher to concentrate the areas for week students. As this system shows a good response and performance from 3 schools the administrators can provide e-learning for other subjects also. The unspecified part of this tutor is that the teacher plays a major role in inserting/updating the problems[6]. Each time a student enters he is going to get new set of problems. So application of EDM methods on student data opens up the possibility for students to develop skills in monitoring their own learning and to see directly how their effort improves their success. Teachers get views into students' performance that assist them adapt their teaching or commence tutoring, tailored assignments, and the like. Using the data, administrators can make policies, execute programs, and adapt the policies and programs to progress teaching, learning, and completion/retention/ rates.

References

[1] Baker, R. S. J. d. 2011. "Data Mining for Education." In International Encyclopedia of Education, 3rd ed., edited by B. McGaw, P. Peterson, and E. Baker. Oxford, UK: Elsevier.

[2] Baker, R. S. J. D., and K. Yacef. 2009. "The State of Educational Data Mining in 2009: A Review and Future Visions." Journal of Educational Data Mining 1 (1): 3–17.

[3] Hamilton, L., R. Halverson, S. Jackson, E. Mandinach, J. Supovitz, and J. Wayman. 2009. Using Student Achievement Data to Support Instructional Decision Making (NCEE 2009-4067). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

[4] Corbett, A. T., and J. R. Anderson. 1994. "Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge." User Modeling and User-Adapted Interaction 4 (4): 253–278. doi: 10.1007/BFO1099821.

[5] Feng, M., N. T. Heffernan, and K. R. Koedinger. 2009. "User Modeling and User-Adapted Interaction: Addressing the Assessment Challenge in an Online System That Tutors as It Assesses." The Journal of Personalization Research (UMUAI journal) 19 (3): 243–266.

[6] Prabha, S.Lakshmi; Shanavas, A.R.Mohamed, "Implementation of E-Learning Package for Mensuration-A Branch of Mathematics," Computing and Communication Technologies (WCCCT), 2014 World Congress on , vol., no., pp.219,221, Feb. 27 2014-March 1 2014,doi:10.1109/WCCCT.2014.37, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6755144&isnumber=6755083

[7] Köck, M., and A. Paramythis. 2011. "Activity Sequence Modeling and Dynamic Clustering for Personalized E-Learning. Journal of User Modeling and User-Adapted Interaction 21 (1-2): 51–97.

[8] Hilbert T. S., Renkl, A., Schworm S., Kessler S., Reiss, K., "Learning to teach with worked-out examples: A computer-based learning environment for teachers," Journal of Computer-Assisted Learning, Vol. 24, 2008, pp.316-332, DOI: 10.1111/j.1365-2729.2007.00266.x.

[9] Siemens, G. Connectivism. "A learning theory for the digital age," International Journal of Instructional Technology & Distance Learning, Vol. 2, No. 1, 2005, pp. 3–7, doi:http://www.itdl.org/Journal/Jan_05/article01.htm.

[10] Scotty D. Craig, Michelene T. H. Chi, Kurt VanLehn, " Improving Classroom Learning by Collaboratively Observing Human Tutoring Videos While Problem Solving," Journal of Educational Psychology, Vol. 101, No. 4, 2009, pp. 779–789, doi: 10.1037/a0016601.

[11] Mohammad Issack Santally,Alain Senteni, "Reconceptualisation of the Teaching and Learning Process through Computer-Mediated Frameworks," International Journal of Instructional Technology and Distance Learning, Vol.2,No.1,2005, pp.23–38, doi:http://www.itdl.org/Journal/Jan_05/article03.htm.

[12] Amershi, S., and C. Conati. 2009. "Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments." Journal of Educational Data Mining 1 (1): 18–71.

[13] K. VanLehn," The behavior of tutoring systems," International Journal on Artificial Intelligence in Education, Vol. 16.No.3, 2006, pp. 227–265, doi: http://dl.acm.org/citation.cfm?

[14] Jeong, H., and G. Biswas. 2008. "Mining Student Behavior Models in Learning-by-Teaching Environments." In Proceedings of the 1st International Conference on Educational Data Mining, Montréal, Québec, Canada,127–136.

[15] Baker, R. S., A. T. Corbett, K. R. Koedinger, and A. Z. Wagner. 2004. "Off-Task Behavior in the Cognitive Tutor Classroom: When Students 'Game the System.'" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). New York, NY: Association for Computing Machinery, 383–390.

[16] Romero, C.,&Ventura,S.(2010),Educational data mining: A review of the state of the art,IEEE Transactions on systems man and Cybernetics Part C.Applications and review, 40(6),601-618.