Educational Data Mining and Learning Analytics for Enhancement and Standardization of Online Education Platform a Review Brief

M Munshi*

***HOD, Computer Science & Engineering, Government Polytechnic, Barwani**

Abstract

In this paper we have presented review of online learning and Educational Data Mining (EDM).We discuss profile of researches in the Educational data mining in context of online learning and current trends and innovation in this field. We also discuss different research category and diversity in Educational data mining and methods and algorithm used to address different types of research problems. Related work and research scope is also summarized to encourage researchers about this topic which is relatively neglected. Ultimate goal of this study is to identify suitable educational data mining methods/algorithms in context of E-VANI (A-VIEW based online learning system).

Introduction:

Online education is the biggest innovation in last 100 years in the field of education. World top universities are offering on line courses. Director technical education M.P also launched an

online teaching program E VANI using online learning platform A-VIEW to conduct classes from University level to polytechnics all over state. Online platforms generate huge amount of data during execution. Educational data mining methods can be used to analyze this data to evaluate performance and give feedback to stack holders for constantly improving the system.

Online Learning Perspective:

To address the shortage of infrastructure and competent teachers which is a big problem especially in rural India online learning is an excellent solution. It is very economic as no extra costs for campus, maintenance, transportation etc needed. It will be the only unit costs of interface and connectivity. To address shortage of competent faculty Director Technical Education launched project E VANI to organize distant education classes from S V Polytechnic Bhopal to its affiliated polytechnic colleges all over state. E VANI is virtual class room solution for online teaching using A VIEW interface. A VIEW provides different communication means like Audio/video interactive whiteboard collaboration etc. A-VIEW log data in files and databases during session at various levels candidate level, class level, college level entire state level generated from keystroke to resources accessed in nested fashion. Other features recorded may be time, any particular sequence and context. All this Data can be aggregated over large number of students and EDM algorithms can explore to build model to answer the questions like

- What sequence strategy is more effective?
- What actions are more effective for learning resulting in higher course grades?
- What action resulted in more student's satisfaction, involvement, progress?
- Which virtual class room solution is appropriate in context?
- Which online resources and pedagogical planning are more efficient?
- And many more such questions and test various educational theories.

Data Mining:

Now days almost in every field online transaction take place. During transaction huge amount of data is generated. Which can be collected and accumulated for analysis. Useful knowledge can

generated from this data. Data mining is a process specially designed to analyze data and generate interesting knowledge from it.

Data Mining is an analytic process to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. The real aim of data mining is to extract pattern and predict.

Data Mining is combination of statistics, AI (artificial intelligence) machine learning and DBMS. An important general difference in between Data Mining and traditional statistical methods is that Data Mining is more data reach approach. Data Mining is not intended to find exact relation between variables, establishing multivariate dependencies between them or finding the underlying functions. Rather main focus is on generating prediction and patterns out of data.

Applications of Data Mining:

Data mining is relatively new technique. It is finding application diverse area. It has been deployed in large scale real world problem in science and business, E commerce, medical and insurance etc. successfully.

- In astronomical physics SKICAT, is a notable system used to analyze images. For cataloging sky objects. This system outperformed traditional computing of classifying faint sky objects.
- In business, main application areas are market pattern analysis, investment forecast, fraud transaction findings, analyze demand of online purchasing etc. Well known marketing application is market basket analysis system (Agrawal 1996).If one buy an item he will buy B and C also. Such predications are very useful for retail shop keepers.
- Crime Detection: Crime branch of police often use this technique to Study pattern of movement of culprits so prevention can be taken .Example if criminal is unknown but large cluster of cases points to one offender then he may be subjected to examination his movement may be consider suspicious.
- Medical and Healthcare. Hospitals, insurance companies government accumulate large amount of data to find which procedure, protocol are better. Which preventive measures may avoid epidemic or which medicine will be required in bulk in particular season etc.

- Education data mining and learning analytics are used to research and build models based on the data stored by LMS and E learning tools during online interaction.

Educational Data Mining:

Educational Data Mining (EDM) is one of the Data mining application. EDM is technique to educational data to analyze it in order to resolve educational research issues.

Educational data mining is emerging as a research area with a suite of computational and psychological methods and research approaches for understanding how students learn. New computer-supported interactive learning methods and tools—intelligent tutoring systems, simulations, games—have opened up opportunities to collect and analyze student data, to discover patterns and trends in those data, and to make new discoveries and test hypotheses about how students learn. Data collected from online learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model building.

Educational data mining researchers (e.g., Baker 2011; Baker and Yacef 2009) 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, motivation, meta cognition, 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 scientific knowledge about learning and learners through building computational models that incorporate models of the student, the domain, and the software's pedagogy.

To accomplish these four goals, educational data mining researchers uses the five categories of technical methods (Baker 2011) described below.

1. Prediction entails developing a model that can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). Examples of using prediction include detecting such student behaviors as when they are gaming the system, engaging in off-task behavior, or failing to answer a question correctly despite having a skill. Predictive models have been used for understanding what behaviors in an online learning environment participation in discussion forums, taking practice tests and the like will predict which students might fail a class. Prediction shows promise in developing domain models, such as connecting procedures or facts with the specific sequence and amount of practice items that best teach them, and forecasting and understanding student educational outcomes, such as success on posttests after tutoring (Baker, Gowda, and Corbett 2011).

2. Clustering refers to finding data points that naturally group together and can be used to split a full Data set into categories. Examples of clustering applications are grouping students based on their learning difficulties and interaction patterns, such as how and how much they use tools in a learning management system (Amershi and Conati 2009), and grouping users for purposes of recommending actions and resources to similar users.

Data as varied as online learning resources, student cognitive interviews, and postings in discussion forums can be analyzed using techniques for working with unstructured data to extract characteristics of the data and then clustering the results.

Clustering can be used in any domain that involves classifying, even to determine how much collaboration users exhibit based on postings in discussion forums (Anaya and Boticario 2009).

3. Relationship mining involves discovering relationships between variables in a dataset and encoding them as rules for later use. For example, relationship mining can identify the relationships among products purchased in online shopping (Romero and Ventura 2010).

- Association rule mining can be used for finding student mistakes that co-occur, associating content with user types to build recommendations for content that is likely to be interesting, or for making changes to teaching approaches (e.g., Merceron and Yacef 2010).

These techniques can be used to associate student activity, in a learning management system or discussion forums, with student grades or to investigate such questions as why students' use of practice tests decreases over a semester of study.

Sequential pattern mining builds rules that capture the connections between occurrences of sequential events, for example, finding temporal sequences, such as student mistakes followed by help seeking. This could be used to detect events, such as students regressing to making errors in mechanics when they are writing with more complex and critical thinking techniques, and to analyze interactions in online discussion forums.

Key educational applications of relationship mining include discovery of associations between student performance and course sequences and discovering which pedagogical strategies lead to more effective or robust learning.

This latter area called teaching analytics is of growing importance and is intended to help researchers build automated systems that model how effective teachers operate by mining their use of educational systems.

4. Distillation for human judgment is a technique that involves depicting data in a way that enables a human to quickly identify or classify features of the data. This area of educational data mining improves machine-learning models because humans can identify patterns in, or features of, student learning actions, student behaviors, or data involving collaboration among students.

5. Discovery with models is a technique that involves using a validated model of a phenomenon (Developed through prediction, clustering, or manual knowledge engineering) as a component in further analysis. For example, Jeong and Biswas (2008) built models that categorized student activity from basic behavior data: students' interactions with a game-like learning environment that uses learning by teaching. A sample student activity discerned from the data was "map probing." A model of map probing then was used within a second model of learning strategies and helped researchers study how the strategy varied across different experimental states.

Discovery with models supports discovery of relationships between student behaviors and student characteristics or contextual variables, analysis of research questions across a wide

variety of contexts, and integration of psychometric modeling frameworks into machine-learned models.

Using these techniques, educational data mining researchers can build models to answer such questions as:

•What sequence of topics is most effective for a specific student?

•What student actions are associated with more learning (e.g., higher course grades)?

•What student actions indicate satisfaction, engagement, learning progress, etc.?

•What features of an online learning environment lead to better learning?

•What will predict student success?

Related Work and History:

- Baker and Yacef (2009) defined EDM as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in" (Baker & Yacef, 2009, p. 1).

- The International Educational Data Mining Society was founded in July 2011, by the International Working Group on Educational Data Mining with aim to support collaboration and scientific development in this new discipline, through the organization of the EDM conference series. Website address of EDM society is http://www.educationaldatamining.org/.
- Karen Cator and Bernadette Adams of the U.S. Department of Education, Office of Educational Technology developed a model for 21century learning powered by technology, envisions ways of using data from online learning systems to improve instruction. The article is available on the Web site at www.ed.gov/technology.
- A systematic method for applying data mining techniques to Moodle usage data as established (Cristóbal Romero, Ventura, & García, 2008).

Conclusion:

Educational data mining is emerging research area with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn. A systematic method for applying data mining techniques to A VIEW usage data could be develop using data mining algorithms. Along with multi model analytics. Future research can examine how the adoption of educational data mining and multi model analysis of online educational platform.

References:

- Bashir, S. (2007). Trends in international trade in higher education: Implications and options for developing countries. Washington, DC: The World Bank. Banik
- Aghai, R. (2003). Patterns of virtual collaboration (Doctoral dissertation, University of Technology, Sydney, Australia).
- Eddy, P. L. (2010). Partnerships and collaboration in higher education. New York: John Wiley & Sons.
- Ilemobade, A. A., & Ballim, Y. (2005). Undergraduate engineering training through institutional collaboration in the Southern African region. South African Journal of Higher Education,
- Keats, D., Beebe, M., & Kullenberg, G. (2003) Using the Internet to enable developing country universities to meet the challenges of globalization through collaborative virtual programmes,