EDUCATIONAL DATA MINING AND LEARNING ANALYTICS - CONVALESCING THE TEACHING AND LEARNING PROCESS

Dr V V NARENDRA KUMAR

Computer Science & Engineering Department, St.Mary's Group of Institutions, Hyderabad.

ABSTRACT

The growing use of internet, mobile phones, tablet PCs etc in recent years has enabled people to learn in different ways. Educational Institutions now have many educational models to choose from e-learning and blended learning (use of books and online learning). This increasing opportunities of online learning for the students and teaching models for instructors brings many challenges in the teaching learning processes. Cyber learning generates huge amounts of data related to learning and teaching processes. This data can be mined and valuable information can be mined so as to improve the students’ performance.

Keywords: Data mining, learning analytics, Educational data mining, cyber learning

1. The Potential of Learning Analytics and Educational Data mining

The potential of analytics and data mining—methodologies that extract useful and actionable information from large datasets has become an important trend in education. When applied to education, these methodologies are referred to as learning analytics (LA) and educational data mining (EDM)[1].

In Educational data mining (EDM) we use various analytical techniques to better understanding of relationships, structure, patterns, and causal pathways within complex datasets. Learning Analytics (LA) emphasises on investigating automatically collected data in a teaching and learning process. Cyber learning emphasises on the integration of learning sciences theories with EDM and LA techniques for a better design of teaching and learning systems.

Now a day’s educational systems are so designed such that they capture and store data of the users when they interact with the system. The captured data is then (e.g., big data, system log data, trace data) analyzed using statistical, machine learning, and data mining techniques for the following purpose:

1. To predict students future by observing their knowledge, behaviour, motivation, and attitudes.
2. To identify productive pedagogical sequences, and suggest how these sequences might be personalized to the students needs and also to study the effects of varied pedagogical enhancements on student learning.
3. To build models of learning processes that incorporate data about students, teachers, understanding of subject matter, pedagogies and principles from learning sciences.
4. To understand collaboration in formal and informal learning environments and design adaptive, data-rich learning systems.

In online learning (e-learning/virtual learning) huge data can be generated through which the learner's behaviour can be adjudged. This will help us to improve and personalize education. New tools and algorithms can be developed through EDM for discovering data patterns and Learning Analytics will help us on applying
these tools and techniques in instructional systems.

Educational Data mining can be used to find sequence of subjects & topics that are most effective to students, activities associated with better learning and securing higher grades, actions that provide satisfaction and commitment and to know the features of an online learning environment which lead to better learning.

Learning Analytics will help us to know whether the students understood a topic better and are when they interested in a new one, whether a student can complete a course in a given time frame and whether a student can achieve a better grade in a specific course with or without the help of a counsellor.

Learning systems have the ability to interact with a student and can deliver personalised subject content and assessments based on their standards and interest. Data stored in the database can be used to judge the student's performance and also to take his feedback of the learning system. His performance and feedback can be visualised through a dashboard. Automated feedback can be used for a diversity of applications like: effective help-seeking strategies, effective problem solving and in giving constructive response.

Educational data mining and learning analytics research[2] are beginning to answer increasingly complex questions about what a student knows and whether a student is engaged. Researchers have experimented with new techniques for model building and also with new kinds of learning system data that have shown promise for predicting student outcomes. Broad areas of applications that are found in practice, especially in emerging companies are as follows:

1. modelling of user knowledge, behaviour and experience
2. user profiling
3. modelling of key concepts in a domain and modelling a domain’s knowledge components
4. trend analysis
5. and personalize the user’s experience.

2. Application areas and Sources of Data

Each of these application areas uses different sources of data[3]. The table 1 gives application areas, queries and the data types needed for analysis.
<table>
<thead>
<tr>
<th>Application Area</th>
<th>Questions</th>
<th>Type of Data Needed for Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>User knowledge modelling</td>
<td>What content does a student know (e.g., specific skills and concepts or procedural knowledge and higher order thinking skills)</td>
<td>Student’s responses (correct, incorrect, partially correct), time spent before responding to a prompt or question, hints requested, repetitions of wrong answers, and errors made. The skills that a student practiced and total opportunities for practice. Student’s performance level inferred from system work or collected from other sources, such as standardized tests.</td>
</tr>
<tr>
<td>User behaviour modelling</td>
<td>What do patterns of student behaviour mean for their learning? Are students motivated?</td>
<td>Student’s responses (correct, incorrect, partially correct), time spent before responding to a prompt or question, hints requested, repetitions of wrong answers, and errors made. Any changes in the classroom/school context during the investigation period of time.</td>
</tr>
<tr>
<td>User experience modelling</td>
<td>Are users satisfied with their experience?</td>
<td>Response to surveys or questionnaires. Choices, behaviors, or performance in subsequent learning units or courses.</td>
</tr>
<tr>
<td>User profiling</td>
<td>What groups do users cluster into?</td>
<td>Student’s responses (correct, incorrect, partially correct), time spent before responding to a prompt or question, hints requested, repetitions of wrong answers, and errors made.</td>
</tr>
<tr>
<td>Domain modelling</td>
<td>What is the correct level at which to divide topics into modules and how should these modules be sequenced?</td>
<td>Student’s responses (correct, incorrect, partially correct) and performance on modules at different grain sizes compared to an external measure. A domain model taxonomy. Associations among problems and between skills and problems.</td>
</tr>
<tr>
<td>Learning component analysis and instructional principle analysis</td>
<td>Which components are effective at promoting learning? What learning principles work well? How effective are whole curricula?</td>
<td>Student’s responses (correct, incorrect, partially correct) and performance on modules at different levels of detail compared to an external measure. A domain model taxonomy. Association structure among problems and between skills and problems.</td>
</tr>
<tr>
<td>Trend analysis</td>
<td>What changes over time and how?</td>
<td>Varies depending on what information is of interest; typically would need at least three data points longitudinally to be able to discern a trend. Data collected include enrolment records, degrees, completion, student source, and high school data in consecutive years.</td>
</tr>
<tr>
<td>Adaptation and Personalization</td>
<td>What next actions can be suggested for the user? How should the user experience be changed for the next user? How can the user experience be altered, most often in real time?</td>
<td>Varies depending on the actual recommendation given. May need to collect historical data about the user and also related information on the product or service to be recommended. Student’s academic performance record.</td>
</tr>
</tbody>
</table>
2. Data mining Algorithms used in EDM/LA

Various Data mining algorithms can be used in assessing the learning environments, students' performance and also in predicting student's future. Few data mining algorithms (Table 2) that can be utilised in EDM / LA are:

Table 2 . Data mining Algorithms used in EDM/LA

<table>
<thead>
<tr>
<th>Problem/ Objective</th>
<th>Algorithm/Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluating undergraduate student academic performance</td>
<td>Using a combination of DM methods like ANN (Artificial Neural Network), Farthest First method based on k-means clustering and Decision Tree as a classification approach</td>
</tr>
<tr>
<td>To predict the potentiality of students performance who can fail during an online curriculum in a Learning Management System (LMS)</td>
<td>Expectation Maximization, Hierarchical Clustering, Simple k-Means and X-Means as provided in WEKA software has been used.</td>
</tr>
<tr>
<td>Shows the applications of various DM techniques on student academic data.</td>
<td>Apriori Algorithm is applied to academic records of students to obtain the best association rules which help in student profiling – K-means clustering is used to group students categorically.</td>
</tr>
<tr>
<td>To identify the significant variables that affects and influences the performance of undergraduate students</td>
<td>C-Means clustering method</td>
</tr>
<tr>
<td>To develop student profiles of learner behaviour from learner’s activity in an online learning environment and also to create click-stream server data</td>
<td>Two clustering methods used, Hierarchical clustering (Ward’s clustering) and NonHierarchical Clustering method (k-means clustering)</td>
</tr>
<tr>
<td>To analyze the web log data files of a Learning Management System (LMS)</td>
<td>Markov Clustering (MCL) algorithm for clustering the students’ activity and a SimpleKMeans algorithm for clustering the courses</td>
</tr>
<tr>
<td>To predict student’s behaviour in future</td>
<td>UCAM (Unique Clustering with Affinity Measure) algorithm is for clustering which works without giving initial seed and number of clusters</td>
</tr>
<tr>
<td>Deals with clustering of student access patterns or surfing behaviour in an e-learning environment.</td>
<td>Fuzzy sets and Transitive closure</td>
</tr>
<tr>
<td>High dimensional categorical dataset are difficult to cluster therefore a fully automated algorithm has been proposed.</td>
<td>Introduced a Two Phase Clustering (TPC) algorithm that works as follows; First start with an initial partition that contains a single cluster (i.e. the whole dataset) and then continuously try to split the cluster within the partition into two subclusters. Now check the homogeneity of the two sub-clusters. If the homogeneity of the two subclusters is high then discard the initial cluster and add the two sub-clusters to the partition.</td>
</tr>
</tbody>
</table>
How to teach a basic computer skills course to students from rural or urban backgrounds.

<table>
<thead>
<tr>
<th>Applications of EDM methods comprise several steps (Figure 1).</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initially, a design is planned, i.e., the main aim of the study and the required data are identified.</td>
</tr>
<tr>
<td>2. The data is then extracted from the appropriate educational environment.</td>
</tr>
<tr>
<td>3. Since data may come from several sources or have different formats and levels of hierarchy, data needs to be pre-processed very frequently.</td>
</tr>
<tr>
<td>4. Models or patterns have to be interpreted, which are obtained by applying EDM methods.</td>
</tr>
<tr>
<td>5. Apply the above changes to the teaching learning process.</td>
</tr>
<tr>
<td>6. If for any reason such as problem has not been adequately addressed, the raw data are small or not suitable, or the selected methods are not powerful enough it is not convincing then modify the teaching/learning process or the study design.</td>
</tr>
<tr>
<td>7. The analysis is then again repeated with the modified one.</td>
</tr>
</tbody>
</table>

A hierarchical cluster analysis was performed on the questionnaire data to map out the approaches to teaching profiles of teachers in higher education on the basis of their scores on the ATI abbreviated for Approaches to Teaching Inventory (ATI).

Applications of EDM methods comprise several steps (Figure 1).

1. Initially, a design is planned, i.e., the main aim of the study and the required data are identified.
2. The data is then extracted from the appropriate educational environment.
3. Since data may come from several sources or have different formats and levels of hierarchy, data needs to be pre-processed very frequently.
4. Models or patterns have to be interpreted, which are obtained by applying EDM methods.
5. Apply the above changes to the teaching learning process.
6. If for any reason such as problem has not been adequately addressed, the raw data are small or not suitable, or the selected methods are not powerful enough it is not convincing then modify the teaching/learning process or the study design.
7. The analysis is then again repeated with the modified one.

Figure 1 - Overview of EDM methods

A model of online learning system:

1. A student is given an online learning system which contains some courses.
2. The student is then allowed to pick a course of his interest.
3. The course content is given in 3 or more forms like e-books (pdf files), ppts and animated course content.
4. As the student starts learning his experiences are collected in a database.
5. This data is used to predict student’s performance.
6. These predictions and the student’s feedback about the learning system analysed and then these are presented in a visual form.
7. The students are then advised by the teachers, administrators and developers to improve his performance.
8. They can also supply learning material based on their interest and learning pace.

In the above learning model the following data mining techniques can be used to enhance the students performance.

1. Rough set theory - This can be used find indiscernible (similar) data.
2. Association rule mining - This can be used to find the relation between students performance and the course material he is using.
3. Sequential pattern mining - This can be used to find the rules for where a student makes a mistake and what type of help he needs.
4. Prediction - This can be used to detect the students behaviour based on the form of the course content he is receiving.
5. Finally which are the useful patterns that will help to enhance a student's performance.

Multi criteria Satisfaction Analysis (MUSA) combined with data mining techniques can be effectively used in finding out the satisfaction in the existing teaching and learning methods.

These techniques help an educational data mining researchers to build learning models to find:

1. Sequence of topics and learning methods which are most effective for a specific student.
2. Student actions that are associated with better learning.
3. Various features of an online learning environment that lead to better learning and to
4. Analytics that will predict student success.

A prototypical learning system created by educational data mining [4] and the system-level view of learning analytics may contain the following components:

1. A content management, maintenance, and delivery component interacts with students to deliver individualized subject content and assessments to support student learning.
2. A student learning database (or other big data repository) stores time-stamped student input and behaviours captured as students work within the system.
3. A predictive model combines demographic data (from an external student information system) and learning/behavior data from the student learning database to track a student’s progress and make predictions about his or her future behaviours or performance, such as future course outcomes and dropouts.
4. A reporting server uses the output of the predictive model to produce dashboards that provide visible feedback for various users.
5. An adaption engine regulates the content delivery component based on the output of the predictive model to deliver material according to a student’s performance level and interests, thus ensuring continuous learning improvement.
6. An intervention engine allows teachers, administrators, or system developers to intervene and override the automated system to better serve a student’s learning.

3. Issues, Challenges and Limitations in EDM and Learning Analytics

Various issues related to EDM when combined with other learning analytics include:

1. Combining necessary data from different learning systems.
2. Deciding the data validity and interpreting the results in an useful way so as to enhance the learning behaviours.
3. Using the results in effective decision making.
4. Safeguarding the data privacy and ethics etc.
5. Standards that balance the need for data privacy with the need to link student and teacher

Successful Implementation of EDM and LA depends [4], [5] on sufficient resources for using the big data which is stored in remote and local servers. Also expenses incurring with the software services and storage act as a big challenge for EDM and LA. Some institutions are using data mining and analytics to analyse the learning environments but reducing the memory requirements and which are essential for implementing advanced algorithms cannot be solved in the near future.

Apart from big data challenge what data to collect, focusing on the questions to be answered, and making sure that the data align with the questions also pose a great challenge to the experts.

Yet another challenge is lack of data interoperability since data mining and analytics rely on diverse and distributed data. Over time, piecemeal purchases of software can lead to significant decentralization of the source of education data, such as student information systems, teachers online grade books, homework submission systems, and publishers online assignments, homework help, and assessments.

In educational data mining and learning analytics, student’s learning topics or concepts must be based on the student’s interaction with an online learning system. These must be validated by comparing scores on assessments and course grades. Going beyond one dataset to combining multiple sources of data (e.g., multiple tests, both teacher-made and standardized; behavioural assessments; or online behaviour tracking) in order to provide an integrated view of a student’s progress is not a straightforward task.

The introduction of student-centered, project-based learning is a century-old challenge for educators worldwide. We envision that the integration of constructionist pedagogical approaches with EDM will pave a way for a wider adoption of student-centered approaches since this new interdisciplinary subfield could make assessment more feasible in large scale, enable the building of smarter technologies for real-time feedback, streamline and optimize the process of giving feedback to students, and offer researchers deeper insight into the learning processes in constructionist learning environments.

4. EDM/LA and the Learning Sciences: to the future

Educational data mining and learning analytics made contributions to the learning sciences and also in new research areas. The recent trends suggests that this contribution will continue, and even increase in the years to come.

One key trend is that these methods have been applied to an ever-widening range of data sources. Much of the early work in EDM was conducted within intelligent tutoring systems and much of the work in LA began in web-based eLearning and social learning environments. In recent years, this has extended to a wider variety of educational situations, including data from student collaboration around learning resources, science simulations, teacher newsgroups and school district grade data systems.

A second key is the use of EDM methods to answer an expanding range of research questions, in an expanding range of areas like computer games, argumentation, computer-supported collaborative learning, learning in virtual worlds and teacher learning.

As EDM and LA become used in a wider variety of domains, by researchers from a wider variety of disciplines, and within learning systems of a wider variety of types, we will see the potential of these approaches for enhancing both practice and theory in the learning sciences. As this
occurs, there will be opportunities to conduct finer-grained, broader-scale research in the learning sciences, benefiting the field and the learners impacted by developments in the field.

5. Conclusion

Advancements in technology has provided several e-learning environments for the people. Educational Institutes can select several e-learning models to improve the teaching and learning process. These increasing opportunities also pose may challenges to the teachers as well as students. The large amount of data generated can be mined and analysed for new patterns. This valuable information that may be employed to improve students’ performance as well as to fined improved learning/teaching processes. Also several researches are also being done on improving the e-learning process.

References


[4] Marie Bienkowski, Mingyu Feng,Barbara Means(2012), Enhancing Teaching and Learning, Through Educational Data Mining and Learning Analytics: An Issue Brief,