A Review On Research Areas In Educational Data Mining And Learning Analytics

AbhilashaSankari, Dr. ShraddhaMasih, Dr. Maya Ingle

Abstract: Over the last two-decade or so, educational data mining has evolved as an emerging discipline to analyze the type of data that comes from academics. Several research studies has carried out Intelligent Tutoring System (ITS), Difficulty Factor Assessments, Latent Knowledge Estimation, Knowledge Inferences, Recommender System and Social Network Analysis. Gathering evidence of learning from educational setup has laid the foundation of learning analytics and educational data mining. Bayesian Knowledge Tracing (BKT), Q-Metrics, Performance Factor Analysis and Latent Knowledge Estimation methods are useful for the study of student’s success. Other methods like matrix factorization and knowledge components are suited for analyzing the student’s knowledge and performance. On the other hand, knowledge engineering and clustering is useful to develop student models for educational software. The current scope of research areas and methods utilized in educational data mining and learning analytics has discussed in this paper.

Keywords: Educational Data Mining, Learning Analytics

1. BACKGROUND STUDY
EDM is becoming an increasingly important area of research that is apparent through the number of publications on educational data mining in recent times. Two handbooks [1] [2] has published exclusively for EDM research. There are also review papers published in these two areas of research. The first is a review from 1995 to 2005 [3] that classified the EDM research according to data mining techniques used. Second review [4] was improvement over the first. In this, the classification done by the type of data used and based on educational categories defined in [3]. A review of educational data mining regarding clustering algorithms, applicability in educational data mining and their usability [5]. Other present reviews on the classification of tools used in EDM according to the task they perform [6].

2. SOURCES OF DATA FOR RESEARCH IN EDM
Sources of data includes log files, educational computer based systems, student enrollment data, online classes, discussion forum, standardized test and intelligent tutoring systems. Popular websites that provides educational dataset include PSLC Datashop, Peerwise Learning Environment, DataZoa, Github and MOOCs. Although computational and statistical techniques are necessary but they are not sufficient to advance a scientific domain. Then we need some basic understanding of teaching and learning process and requirement of the stakeholder involved in the process such as teachers, students and parents.

Machine learning and data mining techniques provide us expertise in working with massive data. Data mining is a step in the overall KDD (Knowledge Discovery with Database) process that contains preprocessing, data mining and post processing. Data mining is already successful in many domains [7] and now showing practical results in mining educational processes.

Popular Platform of Educational Data Mining
1. DataShop
2. Peerwise
3. RipPLE (Recommendation in Personalized Peer Learning Environment)

<table>
<thead>
<tr>
<th>SNo.</th>
<th>Platform</th>
<th>Reference</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DataShop</td>
<td>[17]</td>
<td>Showed improvement in cognitive models with difficulty factor assessment.</td>
</tr>
<tr>
<td>2</td>
<td>PeerWise</td>
<td>[8]</td>
<td>Showed a comparison of PeerWise and Moodle, and found that PeerWise to be more effective in promoting interaction than Moodle.</td>
</tr>
</tbody>
</table>

Mrs. AbhilashaSankari, Assistant Professor
Department of Computer Science,
St. Aloysius' College, Jabalpur (M.P.)

Dr. ShraddhaMasih, Reader
School of Computer Science & Information Technology, Devi AhilyaVishwavidyalaya, Indore (M.P.)

Dr. Maya Ingle, Professor, School Of Computer Science & Technology, Devi AhilyaVishwavidyalaya, Indore(M.P.)
3. **RIPPLE (Recommendation in Personalized Peer Learning Environment)**

Designed a framework for PeerWise learning environment. It used collaborative filtering techniques upon biased matrix factorization.

**Figure 1: Distinction between educational datamining and learning analytics**

**Educational Data mining and Learning Analytics**

Distinguishing educational data mining with learning analytics, is quiet entwine, since both defined in relatively similar ways. Learning analytics and educational data mining both need similar datasets and research skills with overlapping research areas, but their techniques and methods are different. Learning analytics used techniques such as sentiment analysis, social network analysis, and learner's success predictions for its research. On the other hand, an educational data mining utilizes methods of relationship mining, discovery with models, classification and clustering. The techniques has laid its foundation from automated method for discovery using educational data [12].

**Figure 1:** Distinction between educational datamining and learning analytics

Face-to-face contact is the traditional way of educational system. On the other hand, distance education can provides methods and techniques to access educational programs and lectures from any remote area. In traditional classroom learning environment, monitoring the learners' behavior is straightforward. While in distance educational system, we need some methods to study the student's behavior. Another domain of research in educational data mining is web usage mining, which can be done on adaptive and intelligent web-based educational system (AIWBES). Web usage mining is the study of user's navigation behavior while using the web log files. Several research in the past has been performed on learning management systems. Adaptive and intelligent web-based educational system (AIWBES) is a combination of both, an intelligent tutoring system (ITS) and adaptive educational hypermedia systems (AEHS). Intelligent tutoring systems are computer-based instructional systems that attempt to determine information about a student's learning status, and use that information to dynamically adapt the instruction to fit the student's needs. ITSs, often known as knowledge based tutors, because they have separate knowledge bases for different domain knowledge. The knowledge bases specify what to teach and different instructional strategies specify how to teach.

In ITS research studies, researchers use measures such as gaming the system, off-task behavior and carelessness to analyze the disengagement behavior of students. Engagement on the other hand, detected with three aspects behavioral, affective and cognitive. To detect the off-task behavior of student [15] a machine-learned model that used just log files of student interactions with the tutoring system. Conclusion is that the off-task behavior is the result of student's lack of educational drive, passive aggressiveness and disliking of the subject.

**Figure 2: Categories of distance education**

**Figure 3: Classification of ITS**

---

**Table:**

<table>
<thead>
<tr>
<th>Educational Data Mining</th>
<th>Learning Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationship Mining</td>
<td>Sentiment Analysis</td>
</tr>
<tr>
<td>Classification</td>
<td>Learner's Success Prediction</td>
</tr>
<tr>
<td>Discovery with models</td>
<td>Social Network Analysis</td>
</tr>
<tr>
<td>Clustering</td>
<td>Information Science</td>
</tr>
</tbody>
</table>

---

**Figure 2:** Categories of distance education

***Computer Based Education System***

- Student Enrollment Data
- Log Files
- Online class discussion forums
- Standardized Test
- Intelligent computer tutors

***Educational Data Mining***

- Classification
- Clustering
- Data Mining
- ML

***Learning Analytics***

- Sentiment Analysis
- Learner’s Success Prediction
- Social Network Analysis
- Information Science

### Key Points:

- Improve the quality of course, content, video lectures
- Improvement in educational software
- Identify slack and weak learners and work upon them
- Improve software’s capacity to individualize at writer’s learning experiences.
- Improve the processes of educational system

---

320

**Figure 3:** Classification of ITS
Adaptive nature of AEHS makes it useful for learners. That means learners differ in their learning styles, cultural background, knowledge level and preferences. The adaptive systems make the learning process easier for them, since it can change according to their need and preferences. Rather than applying same approach for all learning, different approaches are for different learners are adopted.

Recommended System for Technology Enhanced Learning [16] is useful for learners to find help on materials or content. RecSystel used filtering techniques followed by RiPLE (Recommendation in Peer-Learning Environment) [10] that used collaborative filtering algorithms. Another area of research under AEHS is difficulty factor assessment. The idea of difficulty factors assessment to presents improvement on Cognitive model.

Educational data mining classifies into five broad categories [13] prediction, clustering, relationship mining, distillation of data for human judgment and discovery with models. Apart from these categories, method like knowledge engineering also exist and works on human judgment.

In prediction, some aspect of the data called predictor variable are used to find some data called the predicted variable. Prediction requires labels. Labels with noise, called bronze standard. Kappa [18] a measure to infer the degree of noise in label. The four type of prediction methods are classification, regression, latent knowledge estimation and density estimation.

Apart from other methods, knowledge engineering works for development of student models. Knowledge Engineering plays a prominent role in reasoning, problem solving and decision-making.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Method</th>
<th>Reference</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge Engineering</td>
<td>[19]</td>
<td>Model gaming the system</td>
</tr>
<tr>
<td>3</td>
<td>Knowledge Engineering</td>
<td>[21]</td>
<td>Student systematically level assessment in interactive simulation</td>
</tr>
<tr>
<td>4</td>
<td>Classification</td>
<td>[22] [23]</td>
<td>Discusses the detectors of student affect such as boredom, joy, frustration, anxiety etc.</td>
</tr>
<tr>
<td>5</td>
<td>Classification</td>
<td>[24]</td>
<td>Develop Detectors of Off-Task behavior</td>
</tr>
</tbody>
</table>

**Table 3: Methods used in educational context through educational data mining and machine learning**

<table>
<thead>
<tr>
<th>S. No</th>
<th>Educational System</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Web-based System</td>
<td>Log files, Association and sequential pattern mining</td>
</tr>
<tr>
<td>2</td>
<td>LMS</td>
<td>Technology Acceptance Model</td>
</tr>
<tr>
<td>3</td>
<td>ITS</td>
<td>Q-Matrix, Machine Learned Models, Latent Response Model</td>
</tr>
<tr>
<td>4</td>
<td>AEHS</td>
<td>Decision Trees</td>
</tr>
</tbody>
</table>

**Tools and Techniques**

Data mining tools and techniques are applied to improve students success prediction, student retention and recommender systems. In recent years, many tools developed for EDM research [1]. These tools fall into four different categories namely statistical and visualization tools, clustering and classification tools, association rule mining and sequential pattern mining tools and text mining.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Tool</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WEKA (Waikato Environment for Knowledge Assessment)</td>
<td>Free software contains large collections of machine learning and data mining algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Simulog (Simulation of User Logfiles)</td>
<td>Tool that simulate log files [26]</td>
</tr>
<tr>
<td>3</td>
<td>ASquare (Author Assistant)</td>
<td>Includes data mining methods[26]</td>
</tr>
<tr>
<td>4</td>
<td>Gismo</td>
<td>Visualization Tool</td>
</tr>
<tr>
<td>5</td>
<td>CourseVis</td>
<td>Visualization Tool</td>
</tr>
<tr>
<td>6</td>
<td>Listen Tool</td>
<td>Visualization Tool</td>
</tr>
<tr>
<td>7</td>
<td>AccessWatch</td>
<td>Analyze web server logs</td>
</tr>
<tr>
<td>8</td>
<td>Gwst</td>
<td>Analyze web server logs</td>
</tr>
<tr>
<td>9</td>
<td>Webstat</td>
<td>Analyze web server logs</td>
</tr>
<tr>
<td>10</td>
<td>Multistar</td>
<td>Association and Classification</td>
</tr>
<tr>
<td>11</td>
<td>Synergo</td>
<td>Visualization Tool</td>
</tr>
<tr>
<td>12</td>
<td>CoTAT</td>
<td>Visualization Tool</td>
</tr>
</tbody>
</table>

**CONCLUSION**

In conclusion, this paper has presented various areas of research in educational data mining and learning analytics. The use of EDM and learning analytics is the key to successful inference model of educational data. In addition to basic approaches in EDM the paper presents methods that work effectively with academic setting. Further scope of research in learning analytics and educational data mining includes modelling the behavior of student in different phases such as while learning a skill and prediction of their performances while learning another skill or successful pattern to achieve a particular goal. Finally, finding correlation between student learning performances when they are exposed to different learning models. We hope this paper helps anyone who is interested in educational
BIBLIOGRAPHY


Title: A Study of the Effect of Educational Games on Student Motivation

Abstract: The purpose of this study was to investigate the impact of educational games on student motivation in the classroom. The study involved a randomized controlled trial with a sample of 120 students from a secondary school. The experimental group received educational games as part of their regular curriculum, while the control group followed the traditional teaching method.

Results: The experimental group showed a significant increase in motivation compared to the control group. The improvement in motivation was observed across all age groups, with the highest increase among students in the age group of 12-14 years.

Conclusion: Educational games can be an effective tool to enhance student motivation in the classroom. Further research is needed to explore the long-term effects and scalability of educational games in diverse educational settings.

Keywords: Educational Games, Student Motivation, Classroom Environment.