An Analysis of Educational Data Mining (EDM) in Advanced Education System

Gutlapalli Prabhakara Rao¹, DLS Reddy²

¹,²Computer Science Engineering, Rise Krishna Sai Gandhi Group of Institutions, Ongole, India

Abstract: Educational data mining is a new discipline in research community that applies various tools and techniques of data mining (DM) to explore data in the field of education. This discipline helps to learn and develop models for the growth of education environment. It provides decision makers a better understanding of student learning and the environment setting in as of EDM. It also highlights the opportunities for future research.

Keywords: EDM, DM, PSCL, NCES, PRS, PLE, KDD, KT, ITS, DSS.

1. Introduction

In recent years, the ease and advancement in education domain has created a vast amount of data. As the volume of data increase so do the complexity and relationship underneath the data [1]. Exploring knowledge from large volume of data is the biggest challenge. Data mining which is also known as Knowledge Discovery in Databases (KDD) is a technology used in different disciplines to search for significant relationships among variables in large data sets. Data mining is mainly used in commercial applications. Now a days, the researcher have shown great interest in using data mining applications in the field of education to efficiently manage and extract undiscovered knowledge from the data [2]. The Educational Data Mining community website [3]defines educational data mining as follows: —Educational Data Mining is 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 EDM is an interdisplinary field which inherits the features from various domains such as Psychometrics, artificial Intelligence, Domain driven, Information retrieval, Machine learning, Learning analytics, Databases, Cognitive Psychology and so on. [4].The large repositories of data generated from different sources should be analyzed to fulfill the goals in education. The main objective of EDM viewed by different Researchers as [5][6]-

1)Student Modeling: it is related with the creation of student models that includes student behavior, learning style, performance and environment in which they can develop their skills and solve their problems.

2)Domain Modeling: it is related with the designing of methods, tools and techniques for the growth of particular branch/institution.

3)Learning System: developing the system for studying the effects of educational support. e.g. Pedagogical support.

4)Building the computational models for learning and learners that consist of students, domain.

5)Study the effects of resources related to infrastructure, human resource, and Industry-academic relationship in the organization.

To meet all the above mentioned objectives, a study of EDM is required for delivering the quality education. This objective of this paper is to provide brief knowledge of EDM to the researchers or non expert user in this field. It highlights the different modules of EDM and the opportunities for future research. The paper is organized into Sections. Process of EDM is defined in Section 2. Different phases of EDM are described in Section 3.Future Research Directions are given in Section 4.Section 5 conclude the Introduction.

2. EDM Process

The process of Educational data mining is an iterative, Knowledge discovery process which consists of Hypothesis formulation, Testing and refinement [4] (see Fig-1). Hypothesis is developed from various educational environments. It creates large volume of data. The main process of EDM starts with validating data (i.e. finding relationship between variables/parameters/data items). This is also known as preprocessing of data. After preprocessing various DM techniques, tools will be employed on processed data and final results/interpretation will be given to different user of education. Further recommendation will be suggested for the refinement of problems/task.

![Figure 1: Process of EDM](image-url)
3. EDM Modules

The main modules of EDM are User and Stake holders of Education, Tools, Techniques and Models of DM, Educational Data, Task and Results that will altogether used to achieve the objectives of EDM.

A. Users and Stakeholders: According to researchers there are four main groups of users and stakeholders involved in accomplishing the specific objectives.

1) **Learners (Student/Pupils):** The main objective of Learners is to reply to student requisites, improve learning performance, and provide feedback or recommendations to learners.

2) **Educators (Teachers/Faculties/Tutors):** Their objective is to analyze student behavior, learning, cognitive and social aspect and reflect on their own teaching methods.

3) **Educational Researchers:** The main objective of researchers is to develop new tools and techniques for the growth of educational system.

4) **Administrators:** The main objective is to utilize and enhance available resources (human and material) and their educational offer and so on.

B. EDM Methods: One of the essential modules of EDM is the methods of DM, used for different purpose. Romer Ventura [7] and Ryan Baker [8] categorized the methods as:

- **Prediction:** This technique is used to derive predicted variable (single variable) from predictor variables (combination of variables). Prediction is used analyze student performance and drop out.[9,10] and for detecting student behavior.[11]. It is classified into three types.

  - **Classification:** used to predict class label from (discrete or continue). Some popular classification methods include logistic regression, support vector machines and decision trees [12].

  - **Regression:** used to predict from continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks [13].

  - **Density Estimation:** probability density function is used to predicted variable. Density estimator can be based on variety of kernel functions, including Gaussian function.

- **Clustering:** Clustering is an unsupervised classification process .it is used for grouping objects into classes of similar objects [14]. Data items are partitioned into groups or subsets (clusters) based on their locality and connectivity within N-dimensional space. In educational data mining, clustering has been used to group students according to their learning [15].

- **Relationship mining:** Relationship mining is used to determine relationship between variables in a data set and form rules for specific purpose. Relationship mining is classified into four types:

  - **Association rule mining:** This method is used to identify relationship between attributes in data set, extracting interesting correlations, frequent patterns among data items,[16] for finding students’ mistakes often occurring together while solving exercises [17].

  - **Correlation mining:** This method is used to find Linear correlations between variables (positive or Negative). Correlation analysis is used to find the most strongly correlation attributes.

  - **Sequential pattern mining:** This method is used to find inter-session patterns such as the presence of a set of items followed by another items in a time-ordered set of sessions or episodes.[18] based on temporal relationship between variables to predict which group a learner belongs to. Wang et. al. proposes a four phase learning portfolio mining approach. [19]

  - **Casual data mining:** This method is used to find casual relationship between variables by analyzing the covariance of two events or by using information about how one of the events was trigger.

Other Methods are:

- **Distillation of data for human judgment:** The objective of this method is to present data in summarize and visualized way for e.g. (3D graph etc), to focus on appropriate information and support decision making. In EDM it is used for identification and classification [20].
Discovery with models: This type of model is used as component in other analysis such as relationship mining or prediction [2].

Knowledge Tracing: This method is used to monitor student knowledge and skills over time. It is an effective method in cognitive tutor system [21]

C. EDM Application: Ryan baker [6] has divided the application areas into four major sections such as student Models, Knowledge Domain, Pedagogical support provided by learning software, Scientific discovery about learning and learners. For the development of academic and administrative sections of institutions, there are several applications or tasks that have been resolved through DM techniques.

D. EDM Environment: These are the domains in which different users/stakeholders learn. EDM environments are classified into following three categories.

i) Traditional Class room Environment: It is a formal environment in which users of education communicate directly with each other (i.e. face-to face). For e.g. schools, colleges where lectures are delivered by teacher to students in classrooms. [4]

ii) Online/Web Based Environment: It is an informal environment in which users of education make use of internet. For e.g. e-learning [29], Web Based performance prediction [12].

iii) Computer Based Learning: It is hybrid environment of both (formal and informal interaction). In computer based Environment user can work –

Offline: Intelligent Tutoring System (ITS) [30]. Learning Management [31]. Online for e.g., e-learning [24, 29], Collaborative learning [32].

E. EDM Data: The large volume of data gathered from distributed and diverse fields are used for decision making and learning process in educational context. Data collected depends on the EDM environment discussed in section D.

i) Private data: Direct environment generates offline or private data (related to data collected from academic institution).

ii) Public data: Indirect environment generates online or public data (generally related to e-learning, web logs, e-mail, text data etc. EDM research is more feasible with the advent of public educational data repositories such as Pittsburgh Science of Learning Center DataShop (PSLC), National Center for Education Statistics (NCES) [33, 34].

F. EDM Tools: There are various tools for mining the repositories of data based on their usage, functionality, and working environment. [42]. Different tools will be used on the basis of the respective goal.

4. Future Work

Researchers has share their experience and research opportunities in all aspect of EDM are

- Development of system that reduce instructor intervention. such as (DSS,PRS).
- Developing more generalized tools that can be used by expert and non expert user easily.
- Integration with e-learning.
- Standardization of data and models.

Plagiarism is most concerned topic among research scholar. Thus there should be predictive models to detect plagiarism. As, EDM is a growing field and involved in education technology we should find best ways to discover, learn, utilize technology that can improve learning, teaching and leading in the 21st century. Emphasis should be given on-

- Integrating Pedagogy support and data mining.
- Building model to improve personalized learning. For e.g. finding association between teachers and learners.
- Building tools to enhance Open Education Resource.
- Describing more ways of Flipped Classroom Concept (Model that has changed traditional classroom concept to modern Mobile learning).
- Developing models for building User Centric EContent (i.e. EBook).

5. Conclusion

Educational data Mining (EDM) has been evolved as multidisciplinary scientific learning area, rich in data, methods, tools and techniques used to provide better learning environment for educational users in educational context. This paper integrates all the modules of EDM required to facilitate the objectives of educational research. Lastly it shows that, there are many more research topics that exist in this domain. Utilization of data mining techniques within education environment requires a joint effort by the ICT specialists, educationists and the learners.

References


Author Profile

**Gutlapalli Prabhakara Rao** obtained B.Tech degree in computer science and engineering (CSE) from BVSR engineering college, Chimakurthy. At present pursuing M.Tech in computer science and engineering (CSE) department at RISE Krishna sai Gandhi group of institutions, valluru(v), ongole.

**DLS REDDY** obtained degree in MSc technology from JNTU Hyderabad and MTech in computer science and engineering from sathyabama university, Chennai. He got PhD(CSE) from JNTU, Hyderabad. He has 15 years teaching experience and now working as head of the department (CSE) at rise Krishna sai Gandhi group of institutions, valluru(v), ongole.