

# EDUCATIONAL DATA MINING: A review of the state-of-the- art

Purva Bansal

May 31, 2014

- 1 Introduction
- 2 Data Mining Techniques
- 3 Various tasks
- 4 Analysis and visualization of data
- 5 Providing feedback for supporting instructors
- 6 Recommendations for students
- 7 Detecting undesirable student behaviors
- 8 Planning and scheduling
- 9 References

EDM is concerned with developing methods to explore the unique types of data in educational settings and, using these methods, to better understand students and the settings in which they learn. Various data from the Internet and other educational softwares provides large amount of educational data. This paper has reviewed the various methods that are used.

# Data Mining Techniques

- 1 Classification
- 2 Clustering
- 3 Association Rule Mining
- 4 Sequential Mining
- 5 Text Mining

# Various tasks cum steps involved

- Analysis and visualization of data
- Providing feedback for supporting instructors
- Recommendations for students
- Detecting undesirable student behaviors
- Planning and scheduling

# Analysis and visualization of data

- Objective: highlight useful information and support decision making
- help educators and course administrators to analyze the students course activities and usage information to get a general view of a students learning
- methods used:
  - Statistics: concerning the collection, analysis, interpretation or explanation, and presentation of data tells about the
    - browsers students tend to use
    - patterns of use over time
    - statistical indicators on the learners interactions in forums
    - number of different pages browsed, total time for browsing the different pages
    - and more..
  - Visualization information:

- Visualization information: uses graphic techniques to help people understand and analyze data
  - studies oriented toward visualizing different educational data such as
    - patterns of annual, seasonal, daily and hourly user behavior on online forums
    - tutor-student interaction data from an automated reading tutor
    - statistical graphs about assignments complement, questions admitted, exam score
    - sequence of learning objects and educational trails

# Providing feedback for supporting instructors

- Objective: provide feedback to support course authors/teachers/administrators in decision making (about how to improve students learning, organize instructional resources more efficiently, etc) and enable them to take appropriate proactive and/or remedial action
- Different than that of data analyzing and visualizing tasks as they provide only basic information
- Several DM techniques have been used in this task, association rule mining being the most common
- Association rule mining reveals interesting relationships among variables in large databases and presents them in the form of strong rules according to the different degrees of interest they might present



## Association modeling usage:

- Analyzing learning data and to figure out whether students use resources and possibly whether their use has any (positive) impact on marks
- Determining the relationship between each learning-behavior pattern so that the teacher can promote collaborative learning behavior on the Web
- Finding embedded information, which can be provided to teachers to further analyze, refine or reorganize teaching materials and tests in adaptive learning environments
- Finding interesting relationships between attributes, solution strategies adopted by learners and so on, from a web-based mobile learning system

### Domain Specific Interactive Data Mining usage:

- Finding the relationships between log data and student behavior in an educational hypermedia system
- Learning decomposition and logistic regression to compare the impact of different educational interventions on learning
- Usage data analysis to improve the effectiveness of the learning process in e-learning systems

# Recommendations for students

- Objective: Ability to make recommendations directly to the students with respect to their personalized activities, links to visits, the next task or problem to be done, etc. and also to be able to adapt learning contents, interfaces and sequences to each particular student.
- Data mining techniques used: association rule mining, clustering and sequential pattern mining.
- Sequence/Sequential pattern mining: to discover the relationships between occurrences of sequential events, to find if there exists any specific order in the occurrences
- Association rule mining has been used to recommend online learning activities or shortcuts on a course web site, for recommending relevant discussions to the students

For making recommendations to courseware authors about how to improve adaptive courses , for course recommendation with respect to optimal elective courses

- Clustering has been developed to establish a recommendation model for students in similar situations in the future,
- For grouping web documents using clustering methods in order to personalize e-learning based on maximal frequent item sets
- Providing personalized course material recommendations based on learner ability
- Other DM Techniques used are: neural networks and decision trees to provide adaptive and personalized learning support
- Decision tree analysis to recommend optimal learning sequences to facilitate the students learning process and maximize their learning outcome
- Learning factor transfers and Q-matrixes to generate domain models that will sequence item-types to maximize learning

# Detecting undesirable student behaviors

- Objective: discover/detect those students who have some type of problem or unusual behavior such as: erroneous actions, low motivation, playing games, misuse, cheating, dropping out, academic failure, etc.
- DM techniques used: Classification, clustering
- Classification algorithms: decision tree neural networks, instance-based learning, logistic regression for predicting/preventing student dropout
- Decision trees to identify students with little motivation
- Clustering: Kohonen nets to detect students that cheat in online assessments, outlier detection to uncover atypical student behavior an outlier detection method using Bayesian predictive distribution to detect learners irregular learning

# Planning and scheduling

- Objective: enhance the traditional educational process by planning future courses, helping with student course scheduling, planning resource allocation, helping in the admission and counseling processes, developing curriculum, etc.
- DM Techniques used: mainly association rules
- Association rule mining has been used to provide new, important and therefore demand-oriented impulses for the development of new bachelor and master courses
- Association rule mining and genetic algorithms have been applied to an automatic course scheduling system to produce the course timetables that best suit student and teacher needs