EDM APPLICATIONS AND REPORT ANALYSIS - CASE STUDY

M.Saranyakala
School of Computer Science, Engineering and applications
Bharathidasan University, Trichy
Saranmuthu22@gmail.com

Abstract

Data mining, also known as Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from huge amounts of data. Educational Data mining (EDM) is an emerging field exploring data in educational context by applying different data mining technique and tools. Our objective of this paper is New EDM applications will focus prediction use and engage in data mining tools and activities, making data processing more accessible for all users of EDM. We describe the EDM application as one of the research area of predicting student information of the process in which prediction, classification, clustering and regression and its algorithm.

Keywords: Data mining, Knowledge Discovery, Educational Data mining, Prediction, Classification, Clustering, Regression

1. Introduction

Data Mining is a process of which finds useful patterns from large amount of data. The development of information technology has generated large amount of database and huge data in various areas. The educational data mining is an emerging discipline, concerned with developing methods for exploring unique types of data that come from education setting, and using those methods to better understand students, and the setting which they learning. The EDM data contains Private (offline) data and Public (online) data. The goals of EDM is prediction student's future learning behavior, Discovering or improving domain models, studying the effects of educational support and advancing scientific knowledge about leaning and learners. The EDM contains various environments. They are Traditional classroom (Formal) environment, Online/web based (Informal) environment and computer based learning (Individual and Interaction) environment.

In Indian culture most of the parents are uneducated. The main aim of in Indian government is the quality of education not for quantity. But the day by day the education systems are changed and in the century a huge number of universities are established by the order of UGC. As the numbers of universities are established side by side, each and every day a millennium of students are enrolls across the country. With huge number of higher education aspirants, we believe that data mining technology can help bridging knowledge gap in higher educational systems. The hidden pattern, association, and anomalies that are discovered by data mining techniques from educational data can improve decision making process in higher educational systems. This improvement can bring advantages such as maximizing educational system efficiency, decreasing student's drop-out rate, and increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system process. In this current era we are using the KDD and the data mining tools for extracting the knowledge this knowledge can be used for improving the quality of education. The main approach of EDM is Discovery with models and Distillation of Data for Human Judgment. The decisions tree classifications and prediction is used in this type of applications.

2. Edm methods

One of the Essential Modules of Educational Data Mining is One of the method of Data mining, It is used for Different purpose. The EDM Methods are,

- ➢ Prediction
- ➢ Clustering

- Relationship Mining
- Distillation of data for human judgment
- Discovery with model
- Knowledge Tracking

2.1 Prediction

This technique is used to derive predicted variable from predictor variables. [1]Prediction is used to analyze student performance, dropout and detecting student behavior. It is classified into three types. They are

- Classification
- Regression
- Density Estimation

2.2 Clustering

Clustering is an unsupervised classification process. [2] It is used for grouping objects into classes of similar objects. Data items are partitioned into groups or subsets 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.

2.3 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. They are
- Association Rule mining
- Correlation mining
- Sequential pattern mining
- Casual mining

2.4 Distillation of data for human judgment

The objective of this method is to present data in summarize and visualized way for focus on appropriate information and support decision making. [4]In Educational data mining it is used for identification and classification.

2.5 Discovery with model

This type of model is used as component in other analysis such as relationship mining or prediction [3].

2.6 Knowledge Tracking:

This method is used to monitor student knowledge and skills over time. It is an effective method in cognitive tutor system.

3. EDM APPLICATIONS

Educational data mining is a young research area. The EDM research includes e-learning, adaptive hypermedia, intelligent tutoring systems, web mining, data mining, etc. The applications of data mining in educational system declare mainly pedagogical aspects of the learner and the systems. EDM brings together researchers and practitioners from computer science, education, psychometrics, statistics, psychology, etc. The areas of primary EDM applications are,

- Analysis and visualization of data

- Providing feedback for supporting instructors

- Recommendations for students

- Predicting student performance

- Student modeling

- Detecting undesirable student behaviors

- Grouping students
- Social network analysis
- Developing concept maps
- Constructing courseware
- Planning and scheduling

4. EDM TOOLS

EDM contains several tools for using online and offline. It is used to predicting student performance, Predicting Student behaviors and grouping student.

TABLE 1: Educational Data Mining Tools and Its Goals

Tools	Goals
Weka Tool	Weka Tool is a open source software that developing Machine Learning Task. It is used to several decision tree, Classification and Clustering algorithm.
Moodle Tool	MOODLE is a popular open source LMS. It is Help Users in Course Management system.
Rapid Miner	Rapid Minner is a software. It is Identify Student Behavior model in virtual courses.
KEEL	KEEL is a Knowledge Extraction based on Evolutionary Learning. It is an open source java software tool. It is use to assess the behavior of evolutionary learning and soft computing based techniques.
TADA-ED	TADA-ED stands for Tool for Advanced Data Analysis in education. It is used to identify patterns in student's online exercise.
DataShop	DataShop to store and analyze public data.
Decision Tool	The decision tools suite is an integrated set of programs for risk analysis and decision making. It is used to analyze factors related to success and failure of student.

5. PREDICTING STUDENT PERFORMANCE USING PREDICTION ALGORITHM

Prediction is used to predict the student performance or learning outcomes based on course activities. The educational data mining prediction has been used for student performance, dropout rate, failure rate and identifying student behaviors. In a student performance is one of the most trend topics in education data mining. The main objective of Educational Institution is to provide the best quality of education to students and to improve their behavior. Prediction of student academic performance is helps to teachers to predict about student success and failure in examination. The prediction is like a warning system to identify potential weak students, so that teachers can take an appropriate action towards them. Otherwise, they can give proper advices to prevent failure in the examinations. Predict the student performance and help them to improve their performance. Student prediction also used potential for an automatic online learning environment of school management support system [1]. The predictive models have been proposed in the literature, including neural networks and Bayesian methods. The prediction can be classified into three ways. They are
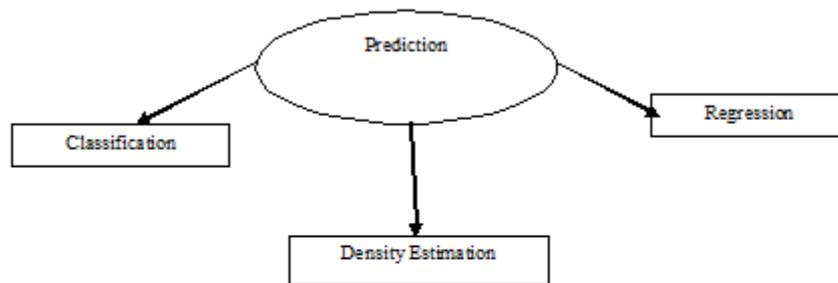
Figure1: Classification of prdiction

Predictive modeling is often a high-level goal of data mining in practice. After outlining the predictive modeling problem, we focus on two classes of algorithm. They are

- Decision tree methods
- Support vector machine (SVM)

5.1 Decision tree methods

Decision trees are especially attractive in data mining environments since human analysts readily comprehend the resulting models. Their construction does not require an analyst to provide input parameters: prior knowledge about the data is also not needed. A record can be associated with a unique leaf node by starting at the root and repeatedly choosing a child node based on the splitting criterion, which evaluates a condition on the input records at the node.

5.2 Support vector machines (SVM)

Support Vector machines are powerful and popular approaches to predictive modeling with success in a number of applications, including hand written digit recognition, charmed quark detection, face detection, and text categorization.

Some Predicting student performance using prediction algorithm is OneR, Ridor, SCart, ADTree, C4.5.

- <u>OneR:</u> In this rule is used to the minimum error attribute for class prediction.
- <u>Ridor:</u> Which is an implementation of the Ripple-Down Rule.
- <u>SCart:</u> SCart is SimpleCart. Which implements minimal cost-Complexity Pruming.
- <u>ADTree:</u> ADtree is an alternate decision. RandomTree, Which considers K randomly chosen attributes at each node of the tree.
- <u>C4.5:</u> C4.5 algorithm for generating a pruned or unpruned decision tree. It is uses Gain Ratio as attribute selection measure to build a decision tree. It's used pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification.

6. PREDICTING STUDENT PERFORMANCE USING CLASSIFICATION

Classification is a form of data analysis that extracts models describing important data classes. Classification has numerous applications, such as fraud detection, target marketing, Student performance prediction, manufacturing, and medical diagnosis. The first steps for classification is to build a model based data. The next step is determined if the model's accuracy is acceptable, and if so, we use the model to classify new data. The educational data mining classification has been used for student performance, dropout rate, failure rate and identifying student behaviors. The main goals of classification are classify student behavior modeling, prediction of student performance, to increase reflection and awareness, Prediction of dropouts. Data mining algorithm such as neural networks and random forests together with decision trees have been applied to predict student's academic success, classifying them into categories of low-risk, medium risk, and high risk of success rate.

Classification methods include decision trees, logistic regression and support vector machines. The regression methods within educational data mining include a linear regression, neural networks, and support vector machine regression. The Classification technique like decision trees, Bayesian networks etc can be used to predicting student performance. Some traditional classification algorithms are,

- Rule-Based Algorithm
- Tree-Based Algorithm
- Function-Based Algorithm
- Bayes-Based Algorithm
- Network Based Algorithm

TABLE 2: Applying some traditional classification algorithm

Bases	Algorithm
Rule Based Algorithm	JRip, NNge, CART, Riod, Rule induction, Ridor.
Tree Based Algorithm	ID3(Iterative Dichotomiser3), C4.5, Decision Tree, ADT(Alternate Decision Tree), LAD Tree, Random Forest, CART(Classification and Regression Trees).
Function Based Algorithm	Logistic, Multilayer Perception, SMO, K-Means.
Bayes Based Algorithm	Bayes net, Naïve Bayes
Network Based Algorithm	Artificial neural network, Genetic Algorithm, RBF network.

6.1 Decision Tree

Decision Tree is a flowchart-like tree structure, each internal node is an attribute, the tree branch node is representing an outcome of the test and the leaf node is a class label. The Tree-shaped structures that represent sets of decisions, these decisions generate rules for classification of a dataset. Specific decision tree methods

include CART-Classification and Regression Trees, and CHAID-Chi Square Automatic Interaction Detection, ID3 and C4.5.

6.2 ID3 Algorithm

- The decision tree each node corresponds to a non-categorical attribute. A leaf of the tree specifies the expected value of the attribute for the records described by the path from the root to that leaf.
- The decision tree at each node should be associated the non-categorical attribute which is most informative among the attributes not yet considered in the path from the root.

6.3 C4.5 Algorithm

The C4.5 algorithm builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. A recursive structure of decision tree is

- A leaf node label with a class value
- A test node that has two or more outcomes, each linked to a sub tree.

6.4 Naïve Bayesian Classification

Naïve Bayesian Classification to be comparable in performance with decision tree and selected neural network classifiers. Bayesian Classifiers have also exhibited high accuracy and speed when applied to large databases.

6.5 Support Vector Machines

Support Vector Machine (SVM), a method for classification of both linear and nonlinear data. SVM is an algorithm works as new dimension and optimal separating hyperplane. The SVM finds hyperplane using support vectors.

6.6 Genetic algorithms

Genetic algorithms attempt to incorporate ideas of natural evolution. The Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

7. PREDICTING STUDENT PERFORMANCE USING CLUSTER

Clustering is a process of grouping objects into groups of similar objects or subsets or clusters. Clustering is a partition a set of records into several groups such that similar records are in the same group according to some similarity function, identifying similar sub populations in the data. Cluster and classification are both classification methods, although clustering is an unsupervised method and classification is a supervised method. The advantage of using classification via clustering is the ability to obtain a general display of the two groups or clusters which are generated are students' performance via activity [5]. Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group of data. Cluster analysis is used to segment a large set of data into subsets called clusters. The clustering methods can be compared many ways. They are the partitioning criteria, separation of Clusters, Similarity measure and Clustering space. The cluster methods are

- Partitioning methods
- Hierarchical methods
- Density-based methods
- Grid-based methods

TABLE 3: Some Cluster Methods

Methods	Characterstics
Partitioning methods	❖ Find mutually exclusive clusters of spherical shape ❖ Distance-based ❖ May use mean or medoid ❖ Effective for small-to medium-size data sets
Hierarchical methods	❖ Clustering is a hierarchical decomposition ❖ Cannot correct erroneous merges or splits ❖ May incorporate other technique like microclustering
Density-based methods	❖ Can find arbitrarily shaped clusters ❖ Clusters are dense regions of objects in space that are separated by low density regions. ❖ Cluster density: Each point must have a minimum number of points within its "neighborhood" ❖ May filter out outliers.
Grid-based methods	❖ Use a multiresolution grid data structure ❖ Fast processing time.

8. Conclusion

In this paper, we include various algorithms to predicting student performance based on prediction, classification and cluster with their supporting tools, available EDM methods for different EDM applications. Each algorithm processed depends on their individual requirement and it varies from one another. So there is need for further implementation for each algorithm based on their future development. The future scope is developing new tool with supporting enormous algorithm to implementing the EDM application. To implement the process may prove same optimal solution by using various algorithm as well as Mathematical computation.

References

[1] Ahamad akardan, Hamid sadeghi, Sep 2013, "A Decision support system for course offering in online higher educational system", *International Journal of computation intelligence system,* vol.6, No.5, 928-942.
[2] Bharahate sachin R,2012, "A survey and Future vision of data mining in educational field", *2nd International conference on advance computing and communication technologies*, IEEE.
[3] Scheuer, M.C. laren, 2011," Educational data mining" In the encyclopedia of science learning, *Springer.*
[4] Osmanbegovic, Sulijic , 2014 "Prediction of students" success by applying Data Mining Algorithm", *Journal of Theoretical and Applied information Technology*, 61(2). 378-388.

[5] S. Vijayakumar and S. Saravanakumar, 2011,"Future Robotics Memory Management", *Advances in Digital Image Processing and Information Technology*, pp.315-325

[6] A.Dinesh Kumar, 2014 " A Survey on predicting student performance", *International Journal of Computer Science and Information Technologies*, Vol.5 (5), 6147-6149.

[7] S. Vijayakumar and S.Saravanakumar, 2011, "Future Robotics Database Management System along with Cloud TPS", *International on cloud computing:Services and Architecture(IJCCSA),pp. 103-114.*

[8] Jasvinder Kumar, 2015, "A Comprehensive Study of Educational data mining", *International Journal of Electrical Electronic and Computer Science Engineering,* ISSN: 2348-2273.

[9] Ryan S.Baker, 2014," Educational Data Mining: An Advance for Intelligent System in Education" IEEE Society.

[10] A.Nancy, Mar 2015, "Energy Efficient Routing Protocols for MANNETs : A Survey ", *International Journal Computer Science and Engineering Technology(IJCSET)*, ISSN:2229-3345,Vol 6,Issue 3.