A Detailed Analysis of Educational Data Mining

Ravi Tiwari¹ and Awadhesh Kumar Sharma²

¹Department Of Computer Science and Engineering, M.M.M.U.T, Gorakhpur, India
²Department Of Computer Science and Engineering, M.M.M.U.T, Gorakhpur, India

¹tiwari.ravi.cs@gmail.com
²akscse@rediffmail.com

ABSTRACT
Educational Data Mining is defined as a discipline which deals with the application of the data mining tools and techniques to explore and analyse the educational data. This discipline helps in the development of the models which facilitate effective decision making from the analysis of the historical educational data. It provides effective solution to various problems faced by the education sector. It provides a direction to the efforts of various entity involves in the education sector which actually leads to efforts utilization. This paper reviews and also describes various data mining techniques and tools used in educational data mining.

Keywords — Educational Data Mining, Learning Analytics, Institutional Effectiveness, Educational Mining Methods, Educational Mining Tools.

1. INTRODUCTION
The educational data mining community [1] defines Educational data mining as, “Educational Data Mining(EDM) 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 setting which they learn in”. Educational Data Mining (EDM) is the application of Data Mining (DM) techniques to educational data, and so, its objective is to analyze these types of data in order to resolve educational research issues. DM can be defined as the process involved in extracting interesting, interpretable, useful and novel information from data [2]. At present time so much data related to the education sector is available in the form file, text, image, video, audio etc. Educational data mining extract useful information from this huge data to develop models which helps in the improvement of the existing learning methods, techniques and environment to facilitate better learning experience.

Educational data has a different nature then the normal data on which traditional Data mining techniques are applied normally therefore new methods and techniques are developed from the traditional data mining techniques which can be directly applied on the educational data. The main goal of the paper is to review the various data mining techniques used in the educational data mining.

2. BACKGROUND AND RELATED WORK
Romero and Ventura [3] done a lot of important work related to education data mining. It uses data mining techniques for the purpose of prediction in education field. Nguyen et al. [4] compares the accuracy of decision tree and Bayesian network algorithm for the prediction of performance of the undergraduate and post graduate students. Results from this work are useful for prediction of the weak student who may fail in the exam. Affendy and Musthpain [5] uses performance in various subjects to predict the CGPA of bachelor students. Al-Radaieh et al.[6] uses classification technique to improve the quality of education. Cesar et al [7] uses data mining technique to develop a model which helps student to take academic decision. Nghe et al also provides the a lot of contribution in this field. Ramaswami and Bhaskaran also develop a predictive model to evaluate achievement of student at higher secondary level. N.S.Shah applies various...

3. EDUCATIONAL DATA MINING METHODS

Various methods used for mining the educational data are as follows-

- Classification
- Prediction
- Clustering
- Regression
- Association rule mining
- Text mining
- Social network analysis

3.1 Classification

Classification methods are used for constructing model or classifier to predict categorical labels such as “pass” or “fail” for the student exam result data , “selected” or “not selected” for the placement data.

It is a twostep process, in the first step first classifier is built describing the predefined set of data, it is known as learning phase. In the second step classifier describe the unknown data into categorical labels. Classification methods can be done with the Naïve Bayes algorithm, decision tree algorithms, neural network algorithms and support vector machines. These methods can be used for classifying student based on their performance, associated risk. It can also be used for the student behaviour modelling.

3.2 Predication

Data predication is also twostep process similar two classifications but it does not use class label terminology because the attribute for which value is predicted is continuous valued instead of categorical .this method is useful in predicting student success rate, dropout rate, and retention rate. Predication can be achieved by naïve bayes algorithm, regression algorithms. Popular regression methods within educational data mining include linear regression, neural networks and support vector machine regression. Many real-world educational data mining problems are not simply prediction. Therefore more complex techniques may be necessary to forecast future values using combination of the techniques (E.g. logistic regression, decision trees or neural networks). For example, the CART (Classification and Regression Tree) Decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables).

3.3 Clustering

Clustering is the process of grouping objects into group based on the similarity. It is actually the process of binding a set of objects into a class based on the common nature they share with each other. Clustering can be achieve through the k-means algorithm, support vector clustering algorithm, random clustering algorithm, agglomerative clustering , top down clustering and DBSCAN clustering algorithm. In the education sector clustering can be used for grouping student based on their behaviour, performance, similarities and for Grouping learning methods based on their effectiveness. It can be used for analyse the learning.

3.4 Regression

Regression is a statistical method of data mining. Regression actually used to model between the one or more dependent variables and independent variables. It can be used for building model or classifier which can analyse the historical data to predict the future trends. In case of educational data this method can be used for student behaviour modelling, for the prediction of the student performance. It can also be used for developing model providing self-awareness.

3.5 Association Rule Mining

Association rule mining is a method to find the frequent item set in the large data set. It is used to find the specific relationship among the data set. Association rule mining discovers relationships among attributes in data set, producing if-then statements concerning attribute-values [3]. The finding of association rule helps in the decision making for example in educational data it suggest may help in the decision making of which teaching methods need to encourage and which is need
to be improved. With respect to educational data it can be used for the student behaviour analysis, student performance analysis, and learning environment analysis. It also can be used for identify student failure, their relationship, their migration and their admission pattern.

3.6 Text Mining

It is defined as the method of extracting useful information from the text. Text mining can be used for extracting information from unstructured data set like emails, documents and files. It can be used for development of the models which can assist students to improve their performance. It is helpful in the automatic construction of text books via text mining. It can also be used for grouping books according to similarity in course and topics. Text mining is also used in evolution of thread discussion [4].

3.7 Social Network Analysis

Social network analysis is a method to measure the interaction and relationships between individual entities in networked information. Social network analysis involves the use of network and graph theories to analyse the social network. The social network analysis techniques and data mining techniques for information networks can be used to examine and assess online interactions [5]. In educational data mining, SNA can be used for mining group activities by analyzing the sociograms associated with a given group and the status of participants and the group cohesion of social interactions [6]. Rallo et al. [7] propose to use the data mining and social networks to interpret and analyse the structure and contents of online educational communities. Social network analysis also helps in student behaviour modelling.

4. EDUCATION DATA MINING TOOLS

The tools used for the purpose of data mining process is also used in the education data mining. Tools used for the education data mining can be classified into categories. First category consist of open source tools table1 and second category table2 consist of commercial education data mining tools.

<table>
<thead>
<tr>
<th>Name of Tool and developer</th>
<th>Functions/Features</th>
<th>Techniques</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrot</td>
<td>Provides ready-to-use components for fetching search results from various sources</td>
<td>Clustering</td>
<td>Windows, Linux</td>
</tr>
<tr>
<td>See5 and C5.0</td>
<td>Provides Decision Tree Analysis, Commercial version of C4.5</td>
<td>Decision Tree</td>
<td>Windows, Unix</td>
</tr>
<tr>
<td>ALPHA MINER</td>
<td>Provides the best cost and performance ratio for data mining applications</td>
<td>Versatile data mining Functions</td>
<td>Windows, Linux</td>
</tr>
<tr>
<td>WEKA</td>
<td>Provides machine learning algorithms for data mining tasks.</td>
<td>Data pre-processing, classification, regression, clustering, association rules</td>
<td>Windows, Linux</td>
</tr>
</tbody>
</table>

Table1: Open source educational data mining tools

<table>
<thead>
<tr>
<th>Name of Tool and developer</th>
<th>Functions/Features</th>
<th>Techniques</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligent Miner (IBM)</td>
<td>Provides tight integration, Scalability of Mining</td>
<td>Association Mining, Classification, Regression, Clustering, Pattern Analysis</td>
<td>Windows, Solaris,</td>
</tr>
<tr>
<td>MSSQL Server (2005)</td>
<td>Provides DM functions</td>
<td>Integrates the algorithms developed</td>
<td>Windows, Linux</td>
</tr>
</tbody>
</table>
Table 2: Commercial educational data mining tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>Provides</th>
<th>by third party</th>
<th>Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mine Set</td>
<td>Provides Robust Graphics tools</td>
<td>Association Mining, Classification,</td>
<td>Windows, Linux</td>
</tr>
<tr>
<td>Oracle Data Mining</td>
<td>Multidimensional data analysis</td>
<td>Association Mining, Classification,</td>
<td>Windows, Mac, Linux</td>
</tr>
</tbody>
</table>

5. CONCLUSION

Educational data is a growing field and have a lot of potential. This field not helps in the improvement of the student performance but it also leads us towards a robust education system which is based on a quick and robust decision making. This field still needs to be exploring to achieve more accurate results. Methods used in educational data mining at present are mainly for the normal business data so there is a need of development of the methods which are according to the nature of the educational data. For the future work intelligent systems can be developed using these data mining methods to mine the educational data.

REFERENCES

[7] Cesar V., Javier B., liela S., recommendation in higher education using data mining, educational data mining conference 2009
[9] Ashkan Sharabiani, member, ieee, fazle karim, anooshiravan sharabiani, mariya atanasov, houshang darabi an enhanced bayesian network model for prediction of students’ academic performance in engineering programs, 2014
[10] Tripti mishra, Dr. dharminder kumar, dr. sangeeta gupta, mining student data for performance prediction, fourth international conference on advance computing and communication technologies, 2014