5 EDUCATIONAL DATA MINING – CHALLENGES AND OPPORTUNITIES IN GLOBAL SCENARIO

By   E. Mohana Roopa,
Assistant Professor
Acharya's Bangalore B-School
Bangalore
Ph no: 91-9880161325
E-mail: emroopa@gmail.com

and   Nila A.Chotai,
Assistant Professor
Acharya's Bangalore B-School
Bangalore
Ph no:9740535589
E-mail: nila.chotai@gmail.com

Abstract

Educational data mining is concerned with developing methods that discover knowledge from data that come from educational environment. The data can be collected from historical and operational data reside in the databases of educational institutes . It can also be collected from e-sources which has a vast amount of information used by most institutes. Educational data mining used many techniques such as decision trees, neural networks, K- nearest Neighbor, Naïve Bayes and support vector machines. Using these methods many kinds of knowledge can be discovered such as association rules , classifications , clustering and outlier detection . The discovered knowledge can be used to better understand students' behavior, to assist instructors to improve teaching , to evaluate and improve e-learning systems, to improve curriculums and many other benefits in the global scenario.

Key Words: *neural networks, e-learning, decision trees*

Introduction

Information era has brought out important changes in many fields, including education a Wi - Fi enabled campus to provide uninterrupted access to information, digital libraries that act as store houses of limitless data, e-learning tools that promote self learning and soon. Large range of web based applications support teaching and learning and comprise adaptive features requiring each user to login and access the information . These recorded data can be stored in a database and mined to discover pedagogically relevant information. There are increasing research interests in using data mining in education. This new emerging field, Educational data mining , concern with developing methods that discover knowledge from data that come from educational environments. The data can be collected from historical and operational data reside in the databases of educational institutes. It can also be collected from e-learning systems which has a vast amount of information used by most institutes. Educational data mining used many techniques such as decision trees ,neural networks ,K-nearest Neighbor, Naïve Bayes and support vector machines.

Using these methods many kinds of knowledge can be discovered such as association rules, classifications , clustering and outlier detection . The discovered knowledge can be used to better understand students' behavior, to assist instructors to improve teaching , to evaluate and improve e-

learning systems , to improve curriculums and many other benefits in global scenario.

Now a days, as in all fields of society, information and communication technologies had busted in education system. So is possible to capture and compile all kind of information in a very easy way and at low cost, such as administrative data of schools, high schools or universities registers ,computerized students academic records, e-learning platforms activity record, computer-supported collaborative learning systems , etc to meet the globalization.

Romero's & Ventura's (2007) review summarizes research papers about educational data mining published in journals and conference proceedings over the 10 years (1995-2005) [4]. The paper surveys the application of data mining to traditional educational systems, particularly Web-based courses, Learning Content Management systems , and adaptive and intelligent web-based educational systems.

Each of these systems has different data source and objectives for knowledge discovering. After pre-processing the available data in each case , data mining techniques can be applied : statistics and visualization; clustering ,classification and outlier detection; association rule mining and pattern mining; and text missing.

EDM is an emerging discipline on the intersection of data mining and pedagogy. On one hand , Pedagogy contributes with the intrinsic knowledge of learning process. On the other hand data mining adds the analysis and information modeling techniques . EDM as data mining, is about a technology that integrates multidisciplinary techniques, from database information retrieval through simple SQL queries to the use of automatic learning algorithms coming from Artificial Intelligence field. Among most used techniques on educational field , are clustering and sequences frequent associations and pattern analysis. For example in mining data about how students choose to use educational software, it may be worthwhile to simultaneously consider data at the keystroke level, answer level , session level, student level, classroom level and school level.

Background:

Educational data mining can be categorized as (according to Romero and Ventura [2007]) as

- ➢ Statistics and visualization
- ➢ Web mining

The methods listed above are quite prominent in Educational Data Mining today, both in mining of web data and in mining other forms of educational data.

Another view point on EDM is given by Baker, which classifies work in EDM as follows:

- ➢ Prediction
- ➢ Classification
- ➢ Regression

> ➤ Density estimation
> ➤ Clustering
> ➤ Relationship mining
> ➤ Association rule mining
> ➤ Correlation mining
> ➤ Sequential mining
> ➤ Casual data mining
> ➤ Distillation of data for human judgement

Discovery with models

The Distillation of data for human judgment has had a prominent place both in published EDM research and in theoretical discussions of educational data mining

The last category of Bakers' EDM taxonomy is perhaps the most unusual category. From a classical data mining perspective. In discovery with models, a model of phenomenon is developed through any process that can be validated in some fashion (most commonly, prediction or knowledge engineering), and this model is then used as a component in another analysis , such as prediction or relationship mining . Discovery with models has become as increasingly popular method in EDM research, supporting sophisticated analyses such as which learning material sub-categories of students will most benefit from [Beck and Mostow 2008], how different types of student behavior impact students' learning in different ways [Cocea et al. 2009] , and how variations in intelligent tutor design impact students' behavior over time [Jeong and Biswas 2008].

Educational data mining has emerged as an independent research area in recent years, culminating in 2008 with the establishment of annual International Conference Of Educational Data Mining , and the Journal of Educational Data Mining.

The success of the plentiful work needs much more specialized work in order for educational data mining to become mature area.

Definition:

"Data Mining or Knowledge Discovery in databases (KDD) is the process of discovering hidden patterns and knowledge within large amounts of data and making predictions of individual outcomes and behaviors".

"Educational Data Mining(EDM) is defined as the area of scientific enquiry centered around the development of methods of making discoveries within the unique kind of data that come from educational settings , and using those methods to better understand students and the settings, and using those methods to better understand students and the settings which they learn in.

Objectives:

Educational Data Mining (EDM) objectives could be identified as :

➢ Scientific research that relates to the creation of knowledge.

➢ Teaching that concerns with the transmission of knowledge.

➢ Institutional research that pertains to the use of knowledge for decision making.

➢ Pedagogic objectives (help in the design of didactic contents, improvement on the academic performance of the students).

➢ Management objectives (to optimize the organization and maintenance of educational infrastructures, areas of interest, more requested courses) and

➢ Commercial objectives (it allow to make market segmentation and facilities students' recruitment, which is especially important in the case of private education).

Applications of Educational Data Mining

There have been a wide number of applications of educational data mining, to mention a few :

➢ Improving student models – which provide the information about a student's characteristics or state such as the student's current knowledge, motivation, meta-cognition, and attitudes. Modeling student individual differences in these areas is the key aim in developing an educational software research and enables software to respond to those individual differences , significantly improving student learning. [Corbett 2001].

➢ Expansion in the sophistication of student models - which enabled researchers to make higher-level inferences About students' behavior, such as when a student is gaming the system, when a student has "slipped"(making a error despite knowing a skill), and when a student is engaging in self-explanation(cf. Shih, Koedinger, & Scheines, 2008) . These student models have increased our ability to predict student knowledge and future performance by incorporating models of guessing and slipping into predictions of student future performance which has increased the accuracy of these predictions by upto 48% . These models have also enabled researchers to study what factors lead students to make specific choices in a learning.

➢ Border range - Educational data mining methods have enabled researchers to model a broader range of potentially relevant students attributes in real-time , including higher level constructs than were previously possible. Researchers have also been able to extend student modeling even beyond educational software, towards figuring out what factors are predictive of student failure or non-retention in college courses or in college altogether.

➢ Discovering or improving models of the knowledge structure of the domain- In Educational Data Mining, methods have been created for rapidly discovering accurate domain models directly from data . These methods have generally combined psychometric modeling frameworks with advanced space-searching algorithms, and are generally posed as prediction problems for the purpose of model discovery (for example , attempting to predict whether individual actions will be correct or incorrect, using different domain models ,is one common method for developing these models).

> Studying the pedagogical support provided by learning software. Modern educational software gives a variety of types of pedagogical support is most effective has been a key area of interest for educational data miners. Learning decomposition, a type of relationship mining, fits exponential learning curves to performance data, relating student success to the amount of each type of pedagogical support a student has received(with a weight for each type of support). The weights indicate how effective each type of pedagogical support is for improving learning.

> Scientific discovery about learning and learners. Research on studying whether state or trait factors were better predictors of how much a student would game the system is a prominent method for conducting scientific discovery about learning and learners.

Opportunities:

> Integrating the data mining with pedagogical theories.

> Describing the different ways of improving educational software or teacher support.

> Supporting sophisticated analyses such as learning materials with sub categories by which students will most benefit.

> Computer supported collaborating learning.

> Computer adaptive testing.

> Identifying the variations in intelligent tutor design which impact students' behavior over a period of time.

> Individual learning from educational software .

> To identify behavior of failing students and to warn students at risk before final exam.

> To assist in the development of new curriculum.

> To analyze the student educational information.

> Using dynamic data mining techniques to improve the educational process.

> To have a better perspective on the student progress through the educational processes.

> To analyze the information related to the specifics of the programs , courses and course assignments.

> Identifying and locating details about educational processes that needs improvements, or those that perform very well which could be used as good examples.

> How data mining is applied for understanding learners' needs and personalizing learning?

Challenges

The challenges mainly focused on "what kind of data could be collected", "how could we preprocess the   data", "how to apply data mining methods on data" and finally "how can we benefited from discovered knowledge". The following are the challenges foreseen   towards educational data

mining:

> Processes or methodologies followed to analyze educational data

> Understanding of learners domain representation

> Improvement assessment of learners engagement in the learning tasks

> Impact of different types of student behavior on students learning different ways.

> Factors those are associated with student failure or non-retention in courses

> Focused on preparing their students for their careers

> Provide guidance outside of class

> Obtaining information on social demand and its potential link for college information systems which are vital for college education development.

> Predicting the paths of students and alumni. For example the Institutions would like to know which students will enroll in particular course programs, and which students will need assistance in order to graduate .Are some students more likely to transfer than others? What groups of alumni are most likely to offer pledges.

> Traditional issues such as enrollment management and time-to-degree continue to motivate higher education institutions to search for better solutions

> Establish social opinion collection mechanism

> Issues of time ,sequence, and context play an important role in the study of educational data.

> Reducing complexity of data mining tools for educators

> Standardization of methods and data

> Integration with the e-learning system

> Integration of educational expert knowledge

Conclusion

EDM is the application of data mining in the educational field, with the purpose of obtaining a better comprehension of students learning processes and their global participation on it, directed to an improvement in quality and cost-effectiveness of the education system. Whether educational data is taken from students' use of interactive learning environments , computer-supported collaborative learning or administrative data from schools and universities ,it often has multiple levels of meaningful hierarchy, which often need to be determined by properties in the data itself, rather than in advance. Educational Data Mining is a promising area of research and it has a specific requirements not presented in other domains. The opportunities and challenges of Educational Data Mining in global scenario are illustrated in this paper. It allows having a better insight on the student progress through the educational process, analyzing the information related to the specifics of the programs , courses, and course assignments. And also helps in decision making process to use the what-if scenario when analyzing the student data, and other educational related information in order to improve educational

process.

References

1. Beck, J.E. and Mostow, J. 2008.How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. In proceedings of the 9th International Conference on Intelligent Tutoring Systems, 353-362.

2. Branes, T., Desmarais , M., Romero, C. and Ventura, S.(2009).Educational Data Mining 2009 ; 2nd International Conference on Educational Data Mining ,Proceedings,Cordoba,Spain .

3. Cocea, M., Hershkovitz, A. and Baker ,R.S.J.D. 2009.The impact of Off-task and Gaming Behaviours on learning :Immediate or Aggregate ?. In Proceedings of the 14th International Conference on Artificial Intelligence in Education. 507-514.

4. Corbett, A.T. 2001. Cognitive Computer Tutors :Solving the Two-Sigma Problem. In proceedings of the International Conference on User Modelling ,137-147 .

5. Mostow,J and Beck ,J.,"Some useful tactics to modify, map and mine data from intelligent tutors" .Natural language Engineering 12(2),195-208.2006

6. Romero,C. and Ventura, S., "Educational data mining :A Survey from 1995 to 2005", Expert Systems with Applications(33) 135-146.2007.

7. Romero,C., Ventura , S. and Garcia , E., "Data mining in course management ystems:Moodle case study and tutorial". Computers & Education, Vol.51, No.1.pp 368-384.2008

8. Superby, J.F., Vandamme, J.P. and Meskens influencing the achievement of the first-year university students using data mining methods. In proceedings of the workshop on Educational data mining at the 8th International Conference on Intelligent Tutoring Systems(ITS 2006),37-44.

9. Tanimoto, S.L, 2007. Improving the Prospects for Educational Data Mining. In Proceedings of the complete On-Line Proceedings of the workshop on Data Mining for User Modelling, at the 11th International Conference on user Modelling (UM 2007), 106-110.