

PERSPECTIVES ON EDUCATIONAL DATA MINING: A STUDY

K.G. Nandha Kumar* and A. Jayanthila Devi**

Abstract: Data mining techniques are applied in ample number of fields in the recent decades. Education is one of the domains where the data mining could be applied effectively. The core tasks are classification, clustering and extraction of association rules. These could be carried out with appropriate educational data. From conventional to evolutionary scenario, data mining algorithms, methods, models and techniques are vast. Identifying a particular algorithm for a specific task is important and it is comparatively difficult. This paper represents some well formed methods and techniques of data analysis for education domain. Exploitation of educational data and various kinds of applications are discussed here. In educational data mining, most of the research works are concerned only with predictive analysis and models. There are scopes for other types of researches too. This paper indicates some untouched areas within the educational domain and also discusses some directions to the future research.

Keywords: Data Mining, Education, Assessment, Prediction, Methods and Models.

INTRODUCTION

Data mining techniques are applied in ample number of fields in the recent decades. Particularly in the educational domain, these techniques could be applied to build different kind of phenomena. Data mining is a field of research which has three edges, namely classification, clustering, and rules mining. Data classification is a fundamental task in data mining and analytics field. It classifies the data through predefined data labels. Examples and labels are given before the start of mining process. Hence, it is a supervised method. Data clustering is used to classify raw data, where there are no predefined labels. Labels are given after the end of mining process. Hence it is an unsupervised method. The third task is entirely different. Rule mining or extraction of association rules is a data mining process which searches the data for frequent item sets, associations and sequential patterns. These fundamental tasks are frequently carried out with different data from different fields in order to obtain hidden facts and patterns. The range of application of data mining is uncountable. These could be applied on all kind of electronic data.

Application of data mining methods, algorithms and techniques are more significant for educational research. Therefore educational data mining is considered more significant as an interdisciplinary area of research. Different types of approaches are involved in educational data mining. Normally, a data mining task starts with data collection and followed by preprocessing, data mining

* Assistant Professor, Department of Computer Science, Dr.NGP Arts & Science College, Coimbatore, India

** Assistant Professor, Department of Computer Applications, Jain University, Bangalore, India

(classification, clustering, rule mining), knowledge extraction and visualization of results. Then the mining model can be evaluated during the testing period. There are some conventional techniques for each step of above said process. But in the educational data mining, it is not necessary to apply the conventional techniques and algorithms at all the steps. Manual and statistical methods are also being applied based on the needs.

RELATED WORK

An analysis was made to build a predictive model for students' performance by the authors [1]. Totally 932 records of engineering students of El Bosque University were collected and analyzed by the developed model. The predictive model is built by using Decision Tree algorithm and Artificial Neural Networks. A model is also used to predict slow learners too [2]. By using Decision Tree algorithm and Naïve Bayes a predictive model was constructed to identify loss of academic performance of students. The model generates the list of students who have lost their academic status.

Prediction of school failures and dropouts is possible by educational data mining [3]. In Mexico, 670 middle school students' records were collected and processed by this model. By using Induction Rules and Decision Tree, the model identifies that which student might get fail marks and who will drop out his studies. In addition to that, prediction of instructor's performance is carried out [4]. Decision Tree, Support Vector Machines, and Artificial Neural Networks are applied on students' responses data set. Responses of students were collected through questionnaires.

Informal conversation of students at social media is also used to understand the students' experience [5]. Such type of data can be extracted from social websites such as facebook, twitter, and blogs. A student's performance can be personally predicted through data mining model to identify the ongoing problems of the students [6].

Researchers have used C4.5, ID3 algorithms in the Weka data mining software to predict the dropouts from schools and colleges [7]. Evaluation of learning outcomes and learning context is carried out on students of engineering stream by using a customized model called Parallel Factors Analysis [8]. This is also used to evaluate opinions and performance of students. These kinds of analyses and predictions can be done at higher educational institutions [9] also. Intelligent planner and recommendation system is modeled by using students' profile and interests [10].

Prediction of achievement in a particular subject is another type of research which is carried out on pre-university students [11], [12]. Back Propagation Neural Networks (BPNN), Classification and Regression Tree (CART), Generalized

Regression Neural Networks (GRNN), Multilayer Perceptron (MLP), Naïve Bayes algorithm, rule based and tree based algorithms are used to construct a predictive model. In addition to school, college and university, the same methods can be applied for institutions of distance education as well [13].

Extraction of association rules and clustering techniques were also used to solve the problem of identifying the reasons for low scoring [14]. Questionnaires were used to collect data from the students. Teaching quality of the faculty members were assessed by data mining techniques [15]. Apriori algorithm is frequently used to find association rules and it is also used to find the reason for failures in exams, by using extracted rules [16].

EDUCATIONAL DATA

Nowadays, generation of data in digital format is inevitable. Data bases are filled with huge amount of data by day-to-day activities. Educational institutions are not exempted from this scenario. Hence, it leads to a large collection of educational data. Particularly two third of educational data are about students. Students' feedback, marks, other performance measures, and interests of students can be collected as data sets. Performance based data are regularly generated by organizational softwares. Students' feedbacks can be collected through manual questionnaires or online questionnaires. Students' interests can be collected through interviews, questionnaires, formal and informal conversations with the students. Extraction of opinions of students from social websites is also possible. In the overall picture, student data are found structured, semi-structured, unstructured and mixture of text and other formats.

PREDICTIVE MODELS

In many number of cases, data mining techniques are used to build predictive models in the education domain. While surveying the literature, seven out of ten works are found as predictive models. The researchers focus only on prediction of failures and dropouts in schools, colleges and universities. No such literatures are found, which discuss on prediction of academic performance of very outstanding students and also no works are found about predictions of extra skills of the students. This is a big gap prevailing in the predictive models. Through the predictive models moderate or average students might be identified and their interests could be found through association rules mining. The performance and other type of data of the research scholars are not much focused. Existing literatures represent the application of data mining only on evaluation of performance of the students. They discuss only the final outcomes. No works found about prediction of current problems of students, such as frequent absence to the classes, late submission of assignments, showing

poor interest during the lecture sessions, special interests and lack of interests on other co-curricular activities.

Very few research works are about the prediction of instructor and teachers' performance. Applying such predictive models on mode of distance education are also significant but availability of such literatures are very few, because, attributes and variables of distance education are different from regular education mode. Number of exclusive works is not sufficient in the case of educational data mining in distance education.

ASSESSMENT MODELS

Application of knowledge discovery processes for assessing the students and teachers is highly possible with the existing methods and techniques. However, such techniques are applied only for assessment of regular academic performance. There are scopes for multiple types of assessments in the field of education such as psychometric tests, aptitude tests, reasoning tests, creativity tests, and they are not considered by the researchers as platforms for the application of data mining techniques. Cluster analysis and association rule mining could be applied successfully in these areas. Behavioral analyses, identifying multiple talents of students and identifying psychological problems are also possible through data mining methods. Models for above said issues are very rare. Besides assessment models can be built for different stake holders of education. In this modern scenario, assessing the potency and interest of students, parents, teachers, society, and educational service providers are equally significant. Most of the models are built for offline assessment and more online assessment models are to be built by using data mining techniques.

EDUCATIONAL OPINION MINING

Opinion mining is one of the emerging areas in data mining research [17]. Opinions are collected from social websites. Normally, people express their opinions, feelings, reviews, feedbacks and comments through informal words and emoticons. These unstructured data can be used to find the exact opinion of people. This could be possible in the field of education too. Opinions on online video lectures, and feedbacks of students on educational issues can be extracted from discussion forums, tweets, blogs and facebook posts. After careful preprocessing they can be processed for knowledge discovery. Summarization of students' comments on particular lecture/session, polarity classification and prediction from students' opinions (whether a student likes or dislikes something), prediction of usefulness of classes and activities, students' feedback on individual elements of educational activities are other matters and directions for application of opinion mining on educational data.

CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

In education domain, applications of classification techniques are more when compared with other data mining techniques. Most of the models are predictive models and less amount are assessment models. Tree based algorithms are frequently used in classification and prediction tasks. In some cases artificial neural networks are applied and particularly regular perceptron networks with back propagation learning is applied. Most of the mining tasks are of performance predictions such as poor learners, failures and dropouts. Cluster analysis techniques and association rule mining algorithms are not frequently used. The research directions are not fully understood and not expanded. Very less number of methods and algorithms for accomplishing data mining tasks is used. That means many researchers use habitual existing algorithms. Evolutionary optimization algorithms, machine learning algorithms, and genetic algorithms could be applied in addition to the traditional data mining algorithms. Other supervised and unsupervised algorithms and approaches could be adapted in order to obtain better results. E-learning and E-assessment problems could be considered to apply different data mining paradigms. Focus on automatic assessment and prediction by using artificial intelligence and expert systems to be increased in the educational data mining context.

References

- Merchan Rubiano S.M. and Duarte Garcia J.A.: Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance. IEEE Latin America Transactions. Vol. 14, No. 6, pp. 2783-2788, (2016).
- López Guarín C.E., Guzmán E.L. and González F.A.: A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje. Vol. 10, No. 3, pp. 119-125, (2015).
- Marquez-Vera C., Morales C.R. and Soto S.V.: Predicting School Failure and Dropout by Using Data Mining Techniques. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje. Vol. 8, No. 1, pp. 7-14, (2013).
- Agaoglu M.: Predicting Instructor Performance Using Data Mining Techniques in Higher Education. IEEE Access. Vol. 4, pp. 2379-2387, (2016).
- Chen X., Vorvoreanu M. and Madhavan K.: Mining Social Media Data for Understanding Students' Learning Experiences. IEEE Transactions on Learning Technologies. Vol. 7, No. 3, pp. 246-259, (2014).
- Elbadrawy A., Polyzou A., Ren Z., Sweeney M., Karypis G. and Rangwala H.: Predicting Student Performance Using Personalized Analytics. Computer, Vol. 49, No. 4, pp. 61-69, (2016).
- Heredia D., Amaya Y. and Barrientos E.: Student Dropout Predictive Model Using Data Mining Techniques, IEEE Latin America Transactions, Vol. 13, No. 9, pp. 3127-3134, (2015).
- Fuzzy-Ant Colony based Routing on Road Networks, Published in APTIKOM Journal on Computer Science and Information Technologies Vol. 1, No. 1, 2016, pp. 33~40 ISSN: 2528-2417.

- Rajni J. and Malaya D. B.: Predictive Analytics in a Higher Education Context. *IT Professional*, Vol. 17, No. 4, pp. 24-33, (2015).
- Garrido A. and Morales L.: E-Learning and Intelligent Planning: Improving Content Personalization. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*. Vol. 9, No. 1, pp. 1-7, (2014).
- Chun-Teck Lye, Lik-Neo Ng, Mohd Daud Hassan, Wei-Wei Goh, Check-Yee Law, and Noradzilah Ismail.: Predicting Pre-university Students' Mathematics Achievement. *Procedia - Social and Behavioral Sciences*. Vol. 8, pp. 299-306, (2010).
- Parneet Kaura, Manpreet Singhb, and Gurpreet Singh Josanc.: Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia - Computer Science*. Vol. 57, pp. 500-508, (2015).
- Baha Sena and Emine Ucarb.: Evaluating The Achievements of Computer Engineering Department of Distance Education Students with Data Mining Methods. *Procedia Technology*. Vol. 1, pp. 262-267, (2012).
- Monolis Chalaris, Ioannis Chalaris, Christos Skourlas and Anastasios Tsolakidis. Extraction of rules based on students' questionnaire. *Procedia – Social and Behavioral Sciences*. Vol. 73, pp. 510-517, (2013).
- Wu Yan.: Application Research of Data Mining Technology about Teaching Quality Assessment in Colleges and Universities. *Procedia - Engineering*. Vol. 15, pp. 4241-4245, (2011).
- “Green Technologies for the Energy-optimized Clouds” in *Asian Journal of Research in Social Sciences and Humanities*, Vol. 6, Issue 6, Special Issue June 2016.
- Nandha Kumar K.G. and Christopher T.: Opinion Mining – A survey. *International Journal of Computer Applications*. Vol. 113, No. 2, pp. 15-17, (2015).