

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 7, July 2018

A Study on Educational Data Mining

Kapila Kundu¹, Ramkala Punia²

Research Scholar, Computer Science and Engineering Department, GJUS&T University, Hisar, India¹

Research Scholar, Computer Science and Engineering Department, GJUS&T University, Hisar, India²

ABSTRACT: Data mining can be defined as a technique to find patterns or interesting information in large amount of data. Educational Data Mining (EDM) is an emerging approach which combines data mining and education system. EDM deals with developing and exploring methods for educational data, and using those methods to better understand students' performance, and their learning environment. This paper includes the details of data mining methods and educational data mining application area or tasks. Under data mining methods and educational data mining tasks section, most common methods are included. Classification is most dominating method used in educational data mining. Among the various tasks, predictive modeling is a task that arouses the curiosity of the researchers.

I. INTRODUCTION

Data mining means extracting novel, interesting, useful and comprehensive knowledge from large amount of data [1]. It includes various methods which have been applied in different application area like clustering, classification, association rule mining and sequence mining. Educational Data Mining (EDM) is an emerging approach which combines data mining and education system. EDM deals with developing and exploring methods for educational data, and using those methods to better understand students' performance, and their learning environment [2, 3, 4, 5, 6]. EDM includes a variety of methods that have been presented by the researchers in the field of educational data mining in recent years. Some methods are similar to basic data mining methods like classification, clustering, association rule mining. On the other hand, some methods are unique to educational data mining like regression, ensemble methods [5, 6, 8, 9]. EDM deals with many educational systems like traditional classroom, e-learning, learning management system, adaptive hypermedia educational systems, intelligent tutoring system, test/quizzes, and text/contents [4, 7, 8]. There are numerous application areas or tasks that have been employed in educational environment. Romero (2010), Baker (2012) and Alejandro (2014) enlist many application areas and variety of functionalities related to students, teachers and about organization under the umbrella of EDM [2, 3, 9]. The most commonly used tasks by the researchers are as follows: student modeling, predicting students' behavior, assessment, feedback and support, teacher, curriculum and domain knowledge, and predicting students' performance [9]. These are the frequently used application areas in educational data mining.

II. DATA MINING METHODS

Data mining can be defined as a technique to find patterns or interesting information in large amount of data. There are many methods which are used in data mining like association rule mining, clustering, classification, text mining etc [1]. In this proposal we will focus on classification and clustering.

II.i. ASSOCIATION RULE MINING

Association means relationship or togetherness or connection of objects. Association rule mining indicates associated relationship between the set of objects/items. An example of an association could be that, 90% of the people who buy cookies also buy milk (60% of all grocery shoppers buy both) [1].

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 7, July 2018

II.ii. CLUSTERING

Clustering is unsupervised data mining technique used to classify data into different groups based on their inherent characteristics. Many researchers have applied clustering in student performance modelling, e.g., to divide the students of interest into different groups to look at the common characteristics of the target groups under consideration like the students who have failed, the student who have achieved average score and the students who have done very well [1].

II.iii. CLASSIFICATION

In classification, the predicted variable can either be a binary or categorical variable. Classification is a supervised learning technique which classifies the data on the basis of known class variable. Some popular classification methods include decision trees, random forests, and decision rules. Classifiers are typically validated using cross-validation, where part of the data set is repeatedly and systematically held out and used to test the goodness of the model. Cross-validation should be conducted at multiple levels. It is typically standard to cross-validate at the student level in order to ensure that the model will work for new students. Some common metrics are used for cross-validation in classification i.e. AUC, kappa, precision, recall, and accuracy [1, 12].

II.iv. REGRESSION

Regression deals with continuous data. It is used to find out the relationship between the predicted (dependent) variable and other independent variables. The most popular regression method is linear regression, and logistic regression. Another method, regression tree is also fairly popular. Regressors can be validated using Kappa, linear correlation or root mean squared error (RMSE) [1, 12].

III. EDUCATIONAL DATA MINING TASKS

The tasks are the application areas in educational field. There are various tasks described in the field of educational data mining like predicting student performance, predicting student behaviour etc. we have described six common tasks used frequently in educational environment [9]. These are as follows:

III.i. STUDENT MODELING

Student modeling is used to classify the learner according to his/her characteristics such as emotion, cognition, domain knowledge, learning strategies, achievements, features, learning preferences, skill, evaluation and affects [9, 11, 13, 14, 15, 16, 17].

III.ii. PREDICTING STUDENTS' BEHAVIOR

The main aim of predicting student behavior is to mine particular pattern behavior in order to provide appropriate help to the user [9, 19-23].

III.iii. ASSESSMENTS

Main focus of assessment is to evaluate the learners' domain knowledge, acquisition, skill development, and achieved outcomes, as well as reflection, inquiring, sentiments [25-29].

III.iv. STUDENT SUPPORT AND FEEDBACK

Student support and feedback is used to identify that the support given by the computerized educational system is relevant to enhance the learners' performance and achievements, or to correct their misconceptions, bugs, and faults [9, 30- 33].

III.v. TEACHER SUPPORT, DOMAIN KNOWLEDGE AND CURRICULUM

Teacher support, domain knowledge and curriculum deals with curriculum development, teaching practices, domain knowledge and teacher support [9, 33-36].

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 7, July 2018

III.vi. PREDICTING STUDENTS' PERFORMANCE

Predicting students' performance means to detect the significant attributes, relationship between the attribute, and other aspects that affect students' performance. Many indicators help to detect the students' performance [37-40]. These are as follows: efficiency, evaluation, achievement, competence, resource consuming, elapsed time, correctness, deficiencies, etc. The main target of this task is to assess the learner if he/she will be able to accomplish a given task or reach a specific learning goal [9].

IV. CONCLUSION

Educational Data Mining is very promising approach which extract novel and interesting information from educational data. Classification is most dominating task used under the umbrella of EDM. Predicting students' performance is dominant task in educational data mining that has always aroused the curiosity of researchers. Understanding factors that affect the students' performance is very important facet that helps in improving and understanding the educational landscape. By predicting student's performance, students can be helped at the early stages and this can improve the academic success of students.

REFERENCES

- [1]. J. Han, M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
- [2]. R. S. J. D. Baker, & K. Yacef, "The state of educational data mining in 2009: A review and future vision", *Journal of Educational Data Mining*, vol. 1, issue 1, pp. 1–15, 2009.
- [3]. C. Romero, & S. Ventura, "Educational data mining: a review of the state of the art", *IEEE Transactions on Systems, Man, and Cybernetics, part C: Applications and Reviews*, vol. 40, issue 6, pp. 601–618, 2010.
- [4]. C. Romero, & S. Ventura, "Educational data mining: a survey from 1995 to 2005", *Expert Systems with Applications*, vol. 33, issue 1, pp. 135–146, 2007.
- [5]. C. Romero, S. Ventura, M. Pechenizkiy, & S. J. d. R. Baker (Eds.), "Handbook of educational data mining, data mining and knowledge discovery series", Florida: Chapman & Hall/CRC, 2010.
- [6]. R. S. J. D. Baker & P. S. Inventado, "Educational Data Mining and Learning Analytics", *Learning Analytics: from research to Practice*, pp. 61-75, 2014.
- [7]. O. Zaïane, "Web usage mining for a better web-based learning environment", *In Proceedings of Conference on Advanced Technology for Education*, 2001, pp. 60–64.
- [8]. C. Romero, & S. Ventura, "Data mining in e-learning", book series 4, WIT Press, 2006, pp. 328.
- [9]. A. Pena-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works", *Expert Systems with Application*, vol. 41, pp. 1432-1462, 2014.
- [10]. B. Sen, E. Ucar & D. Delen, "Predicting and analyzing secondary education placement-test score: A data mining approach", *Expert Systems with Applications*, vol. 39, pp. 9468-9476, 2012.
- [11]. S. D'Mello, & A. Graesser, "Mining bodily patterns of affective experience during learning" *In Proceedings of the 3rd international conference on educational data mining*, 2010, pp. 31–40.
- [12]. R. Pelanek, "Metric for evaluation of student model", *Journal of Educational Data Mining*, vol. 7, issue. 2, 2015.
- [13]. R. S. J. D. Baker, S. M. Gowda, M. Wixon, J. Kalka, A. Z. Wagner, A. Salvi, V. Aleven, G. W. Kusbit, J. Ocumpaugh, & L. Rossi, "Towards sensor-free affect detection in cognitive tutor algebra", *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 126–133.
- [14]. H. Guruler, A. Istanbulu, & M. Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases" *Computers & Education*, vol. 55, issue 1, pp. 247-254, 2010.
- [15]. I. Arroyo, H. Mehranian, & B. P. Woolf, "Effort-based tutoring: an empirical approach to intelligent tutoring", *In Proceedings of the 3rd international conference on educational data mining*, 2010, pp. 1–10.
- [16]. J. P. Gonzalez-Brenes, W. Duan, & J. Mostow, "How to classify tutorial dialogue? comparing feature vectors vs. Sequences", *In Proceedings of the 4th international conference on educational data mining*, 2011, pp. 169–178.
- [17]. A. Nandeshwar, T. Menzies, & A. Nelson, "Learning patterns of university student retention", *Expert Systems with Applications*, vol. 38, issue 12, pp. 14984–14996, 2013.
- [18]. O. Montalvo, R. S. J. D. Baker, M. A. Sao-Pedro, A. Nakama, & J. D. Gobert, "Identifying students' inquiry planning using machine learning", *In Proceedings of the 3rd international conference on educational data mining*, 2010, pp. 141–150.
- [19]. M. A. Sao-Pedro, R. S. J. D. Baker, O. Montalvo, A. Nakama, & J. D. Gobert, "Using text replay tagging to produce detectors of systematic experimentation behavior patterns" *In Proceedings of the 3rd international conference on educational data mining*, 2010, pp. 181–190.
- [20]. S. T. Levy, & U. Wilensky, "Mining students' inquiry actions for understanding of complex systems", *Computers & Education*, vol. 56, issue 3, pp. 556–573, 2011.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 7, July 2018

- [21]. Y. Qiu, Y. Qi, H. Lu, Z. A. Pardos, & N. T. Heffernan, "Does time matter? modeling the effect of time in bayesian knowledge tracing", *In Proceedings of the 4th international conference on educational data mining*, 2011, pp. 139–148.
- [22]. J. Bayer, H. Bydzovska, J. Geryk, T. Obsivac, & L. Popelinsky, "Predicting drop-out from social behaviour of students", *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 103–109.
- [23]. J. McCuaig, & J. Baldwin, "Identifying successful learners from interaction behaviour" *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 160–163.
- [24]. P. Kaur, M. Singh, and G. Singh Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector", *Procedia Computer Science*, vol. 57, pp. 500–508, 2015.
- [25]. S. Cetintas, L. Si, Y. P. Xing, D. Zhang, J. Y. Park, & R. Tzur, "A joint probabilistic classification model of relevant and irrelevant sentences in Mathematical word problems", *Journal of Educational Data Mining*, vol. 2, issue 1, pp. 83–101, 2010.
- [26]. K. Wauters, P. Desmet, & W. van den Noortgate, "Monitoring learners' proficiency: weight adaptation in the Elo rating system" *In Proceedings of the 4th international conference on educational data mining*, 2011b, pp. 247–252.
- [27]. M. Feng, N. T. Heffernan, Z. A. Pardos, & C. Heffernan, "Comparison of traditional assessment with dynamic testing in a tutoring system", *In Proceedings of the 4th international conference on educational data mining*, 2011, pp. 295–300.
- [28]. J. D. Gobert, M. A. Sao Pedro, R. S. J. D. Baker, E. Toto, & O. Montalvo, "Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds" *Journal of Educational Data Mining*, vol.4, issue 1, pp. 111–143, 2012.
- [29]. L. A. Sudol, K. Rivers, & T. K. Harris, "Calculating probabilistic distance to solution in a complex problem solving domain", *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 144–147.
- [30]. J. Champaign, & R. Cohen, "An annotations approach to peer tutoring", *In Proceedings of the 3rd international conference on educational data mining*, 2010, pp. 231–240.
- [31]. B. Beheshti, M. C. Desmarais, & R. Naceur, "Methods to find the number of latent skills", *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 81–86.
- [32]. C. Vialardi, J. Chue, J. P. Peche, G. Alvarado, B. Vinatea, J. Estrella, et al., "A data mining approach to guide students through the enrollment process based on academic performance" *User Modeling and User-Adapted Interaction*, vol. 21, issue 1–2, pp. 217–248, 2011.
- [33]. S. Tsuruta, R. Knauf, S. Dohi, T. Kawabe, & Y. Sakurai, "An intelligent system for modeling and supporting academic educational processes" *In A. Pena-Ayala (Ed.), Intelligent and adaptive educational-learning systems: achievements and trends, smart innovation, systems and technologies*, Heidelberg: Springer, 2012, pp. 469–496.
- [34]. M. Chi, K. VanLehn, D. Litman, & P. Jordan, "Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies", *User Modeling and User-Adapted Interaction*, vol. 21, issue 1–2, pp. 137–180, 2011.
- [35]. K. Wauters, P. Desmet, & W. van den Noortgate, "Acquiring item difficulty estimates: a collaborative effort of data and judgment" *In Proceedings of the 4th international conference on educational data mining*, 2011a, pp. 121–127.
- [36]. S. Cai, S. Jain, Y. H. Chang, & J. Kim, "Towards identifying teacher topic interests and expertise within an online social networking site", *In Proceedings of the KDD 2011 workshop: Knowledge discovery in educational data*, 2011, pp. 97–102.
- [37]. Y. H. Wang, & H. C. Liao, "Data mining for adaptive learning in a TESL-based e-learning system", *Expert Systems with Applications*, vol. 38, issue 6, pp. 6480–6485, 2011.
- [38]. A. Zafra, C. Romero, & S. Ventura, "Multiple instance learning for classifying students in learning management systems", *Expert Systems with Applications*, vol. 38, issue 12, pp. 15020–15031, 2011.
- [39]. Z. A. Pardos, Q. Y. Wang, & S. Trivedi, "The real world significance of performance prediction", *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 192–195.
- [40]. Y. Wang, & J. E. Beck, "Using student modeling to estimate student knowledge retention" *In Proceedings of the 5th international conference on educational data mining*, 2012, pp. 200–203.