The 9th International Conference on Cognitive Science

Educational data mining: A review

Siti Khadijah Mohamad[a], Zaidatun Tasir[a],*

[a] Department of Educational Sciences, Mathematics and Creative Multimedia, Faculty of Education,
Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Johor, Malaysia

Abstract

Data Mining is very useful in the field of education especially when examining students' learning behavior in online learning environment. This is due to the potential of data mining in analyzing and uncovering the hidden information of the data itself which is hard and very time consuming if to be done manually. The purpose of this review is to look into how the data mining was tackled by previous scholars and the latest trends on data mining in educational research. Several limitations of existing research are discussed and some directions for future research are suggested.

Keywords: Algorithm; Data mining; Educational data mining; Elearning; Online interaction

1. Introduction

Data mining, often called knowledge discovery in database (KDD), is known for its powerful role in uncovering hidden information from large volumes of data [1]. Its advantages have landed its application in numerous fields including e-commerce, bioinformatics and lately, within the educational research which commonly known as Educational Data Mining (EDM) [2]. EDM is defined by The Educational Data Mining community website, www.educationaldatamining.org "as an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational setting, and using those methods to better understand students, and the settings which they learn in". EDM often stress with the improvement of student models which denote the student's current knowledge, motivation, metacognition, and attitudes [3].

There were collections of reviewed papers that cover the important aspects of data mining in educational research [3,4,5,6]. The first review was concerned on the application of data mining techniques in educational system from the year 1995 until 2005, where each of the systems reviewed has diverse data source and objectives for knowledge discovering [5]. Another review was about applying data mining techniques to e-learning problems [4]. It also reviewed on the use of e-learning in assessing students' learning performance, students' learning behavior, and evaluation of learning material. Next, there was review conducted on the current trends in EDM and shifts in paper topics over the years [3]. More intensive reviewed on each scholar's most important studies and the type of educational task that they were dealing with can be found in Romero and Ventura's work [6]. This paper,

---

* Corresponding author. Tel.: +60 19 7255786; fax: +60 7 5534884.
E-mail address: p-zaida@utm.my

meanwhile, tends to focus on the use of data mining in an online learning environment and more details review can be found in Section 2.

2. Discussion on Selected Papers

In this section, we present the reviewed of 9 latest studies to which the data mining methods are applied in educational setting ranging from the year 2004 until 2012, as stated in Table 1. Most of the studies were gathered from the conferences, and journals publications. The number of citations that each paper received, as for April 22, 2013, indicates the impact that they makes towards educational data mining researchers and the field itself. We begin the review with the brief explanations on each study. The trends and limitations that each study carries are presented in the next subsection.

Table 1. List of studies that focused on educational data mining

References	Objective	Platform	Data Mining Task	Source of Publication	Number of Citations
[7]	Mining patterns of events in students' teamwork data	TRAC system	Sequential Pattern	Conference	47
[8]	Using data mining for automated chat analysis to understand support inquiry learning processes	Online chat	Classification	Conference	24
[9]	Discovering student preferences in e-learning	E-learning	Prediction	Conference	25
[10]	Mining the student online assessment data	E-learning	Classification, Clustering, Association Rule Analysis	Conference	10
[11]	Examining students' online interaction in a live video streaming environment using data mining and text mining	Live video streaming environment	Clustering	Journal	2
[12]	A complete understanding of disorientation problems in web-based learning	Web-based learning system	Clustering	Journal	3
[13]	A web-based intelligent report e-learning system using data mining techniques	E-learning	Classification	Journal	1
[14]	Mining student data to characterize similar behavior groups in unstructured collaboration spaces	Ars Digita Community System	Clustering	Conference	97
[15]	Clustering and sequential pattern mining of online collaborative learning data	TRAC system	Clustering, Sequential Pattern	Conference	71

2.1. Brief explanation on each study

The first study concerns about the information that distinguishes a group that functioning well and weak based on the electronic traces of their collaboration [7]. The collaboration takes place in the TRAC system and consists of three types of events which eventually reflect the students learning process. The mining involve the use of sequential pattern algorithm in order to find the patterns characterizing some aspects of the teamwork. Next, the study investigates the application of data mining methods to provide learners with real-time adaptive feedback while learning collaboratively [8]. The learner's adaptive feedback behavior can be seen by the image of avatar that

derives from the classification of automated chat. The image then became an indicator of the level of learner's participation in the learning environment. The third article concerns with an adaptive user model which able to deal with the student's preferences on educational materials over time [9]. The decision model was developed based on the Bayesian Network Classifier that represents the learning styles and resources in order to decide if the resource is good for student or not. The model can adapt itself to changes according to the student's preferences. The fourth article centers around the small-scale study based on online assessment data where students received an immediate feedback after answering the test [10]. Here, the authors wanted to know whether, different mining techniques which include the use of clustering, classification and association analysis can affect the individual needs of the students. Next, the study is about students' online patterns of interaction in a live video streaming environment. The use of data and text mining were very helpful in gaining insight from large volume of untapped textual data [11]. The sixth article suggests the new framework for the development of web-based learning system which can decrease learners' disorientation problem [12]. Clustering approaches had been used in order to have clear distinction between the clusters. The seventh article proposes a web-based intelligent report of an e-learning which was design to predict the students' behavior pattern with the help from the data mining techniques [13]. Decision tree and neural network algorithm were used in order to create the data mining model. Next, the study is about the preliminary experiments using clustering to build profiles of user behaviors in unstructured collaboration spaces [14]. The discovered profile is then presented to the teacher for supporting interaction assessment. Finally, the last article is concerns with students working in teams and performed mining on data collected to characterize the work of stronger and weaker students [15]. Clustering technique was used to find clusters of similar teams and similar individual members, while sequential pattern mining was applied to extract sequences of frequent events.

2.2. Algorithms used

Based on the meta-analysis done in Table 1, we can see that the most popular techniques for data mining is clustering [10,11,12,14,15], followed by classification [8,10,13], sequential pattern [7,15], prediction [9], and association rule analysis [10]. Back to the year of 1995 until 2005, the association rule analysis method was frequently applied in most of the studies on educational data mining [5], as it requires less extensive expertise than other methods [16]. However, beginning the year of 2005, this is no longer the trend, as researchers often adopting the use of clustering and classification methods in its analysis. The output of association rule is often too many, most of them non-interesting and difficult to understand for non-expert in data mining [17]. In choosing the appropriate algorithms, researchers must first design the data and align it with the desired output. If they have small-scale study, they can opt for clustering approach since this technique does not require the splitting of data as what needed in classification approach. Besides, researchers can always make comparison with different algorithms for the same dataset as what had been done in [10], and this would definitely be something to look for whether similar results will be achieved by using different approach.

2.3. An absence of real collaborative learning process

Referring to Table 1, most of the researchers tend to develop more complex systems in order to collect and examine students' online interaction which later being analyze using data mining techniques. The systems were developed for specific course and cannot be generalized to other types of studies. This certainly makes it hard for educators to apply them as the scope is far beyond what educators may want to do [5]. Overall, all the studies mentioned in Table 1 are focusing on the technical aspects of e-learning, where they investigated which type of modules or features really influence students in learning collaboratively with the help from data mining analysis. Once the module or feature does not work, they change it, and then add another one, in a hope that it will give some benefits towards learning collaboratively among students.

Now, examining how students make use the system is one way to assess the instructional design in a formative manner and it may shed some useful insights for the educator to improve the instructional materials [18]. What seems to be missing in current research is that, the real aspects of collaborative learning approach, which involve joint intellectual effort among students or between students and teachers [19] and how they really engage and connect with the educational learning theories and strategies in learning. Unfortunately, researchers tend to overlook the discourse that happened within the forum or the chat room and the value that it carries towards making learning

collaborative more efficient. These two aspects- the modules or features of the system and the content discourse often evaluated in separate ways. It is certainly be interesting to find out how well these two views can support each other in learning and this can only be done with a help from data mining analysis, as it was born to tackle this kind of complex analysis.

3. Conclusion: Future Directions

Currently, most of the researches on educational data mining pay great attentions towards the use of e-learning like Moodle, WebCT, Blackboard and some even develop their own tool for the learning purposes. With regard of future research, perhaps we can shift our focus from the e-learning, towards the use of social networking tools like Blog and Facebook since these applications already gained high popularity among students and suitable to be used to engage the students with collaborative learning [20,21]. We might, of course, encounter some problems, like difficulties in gathering the log data since these applications are not able to provide us with the logs of learner activity as compared to other e-learning applications, but then again, this can be encountered by integrating the Google Analytics tool into the blog environment and the log data can be exported later for further analysis using the data mining techniques. We hope that this review will be able to shed some useful insights for researchers and educators in order for educational data mining to become a mature area.

Acknowledgements

References

[1] Witten, I.H. and Frank, E. 1999. *Data Mining:Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kauffman, San Francisco, CA.

[2] Baker, R.S.J.d.: Data Mining for Education. In: McGaw, B., Peterson, P., Baker, E. (eds.) To appear in *International Encyclopedia of Education*, 3rd edn. Elsevier, Oxford (2010)

[3] Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.

[4] Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In *Evolution of teaching and learning paradigms in intelligent environment* (pp. 183-221). Springer Berlin Heidelberg.

[5] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.

[6] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. Systems, *Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601-618.

[7] Kay, J., Maisonneuve, N., Yacef, K., & Zaïane, O. (2006). Mining patterns of events in students' teamwork data. In *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems* (ITS 2006) (pp. 45-52)

[8] Anjewierden, A., Kolloffel, B., & Hulshof, C. (2007). Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes. In *International Workshop on Applying Data Mining in e-Learning* (ADML 2007).

[9] Carmona, C., Castillo, G., & Millán, E. (2007). Discovering student preferences in e-learning. In *Proceedings of the international workshop on applying data mining in e-learning* (pp. 23-33).

[10] Pechenizkiy, M., Calders, T., Vasilyeva, E., & De Bra, P. (2008). Mining the student assessment data: Lessons drawn from a small scale case study. *Educational Data Mining* 2008, 187.

[11] He, W. (2012). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*.

[12] Shih, Y. C., Huang, P. R., Hsu, Y. C., & Chen, S. Y. (2012). A complete understanding of disorientation problems in web-based learning. *The Turkish Online Journal of Educational Technology*, 11(3).

[13] Blagojević, M., & Micić, Ž. (2012). A web-based intelligent report e-learning system using data mining techniques. *Computers & Electrical Engineering*.

[14] Talavera, L., & Gaudioso, E. (2004). Mining student data to characterize similar behavior groups in unstructured collaboration spaces. In *Proceedings of the Artificial Intelligence in Computer Supported Collaborative Learning Workshop* at the ECAI 2004.

[15] Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(6), 759-772.

[16] Merceron, A., & Yacef, K. (2007). Revisiting interestingness of strong symmetric association rules in educational data. In Proc. *of Int. Workshop on Applying Data Mining in e-Learning, Creete*, Greece (pp. 3-12).

[17] García, E., Romero, C., Ventura, S., & Calders, T. (2007, September). Drawbacks and solutions of applying association rule mining in learning management systems. In *Proceedings of the International Workshop on Applying Data Mining in e-Learning* (ADML 2007), Crete, Greece (pp. 13-22).

[18] Ingram, A. L. (2000). Using web server logs in evaluating instructional web sites. *Journal of Educational Technology Systems*, 28(2), 137-158.

[19] Smith, B. L., & MacGregor, J. T. (1992). What Is Collaborative Learning?. National Center on Postsecondary Teaching, Learning, and Assessment at Pennsylvania State University

[20] Churchill, D. (2009). Educational applications of Web 2.0: Using blogs to support teaching and learning. *British Journal of Educational Technology*, 40(1), 179-183.

[21] DiVall, M. V., & Kirwin, J. L. (2012). Using Facebook to facilitate course-related discussion between students and faculty members. *American Journal of Pharmaceutical Education*, 76(2).