Predicting Performance Based on the Analysis of Reading Behavior: A Data Challenge

Brendan Flanagan
Kyoto University, Japan
flanagan.brendanjohn.4n@kyoto-u.ac.jp

Atsushi Shimada
Kyushu University, Japan

Stephen Yang
National Central University, Taiwan

Bae-Ling Chen, Yang-Chia Shih
Asia University, Taiwan

Hiroaki Ogata
Kyoto University, Japan

ABSTRACT: As the adoption of digital learning materials in modern education systems is increasing, the analysis of reading behavior and their effect on student performance gains attention. The main motivation of this workshop is to foster research into the analysis of students' interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. In this workshop, participants analyzed the event logs from three different universities datasets with information on over 1000 students reading behaviors. Additional information on lecture schedules were also provided to enable the analysis of learning context for further insights into the preview, in-class, and review reading strategies that learners employ. Finally, workshop contributors were encouraged to implement their research results as a feature of an open LA dashboard.

Keywords: Student Performance Prediction, Data Challenge, Reading Behavior

1   WORKSHOP BACKGROUND

Digital learning materials especially digital textbooks are a core part of modern education, and the adoption of digital textbooks in education is increasing. Digital textbooks and e-books are being introduced into education at the government level in a number of countries in Asia (Ogata et al., 2015). This has prompted research into not only the use of such materials within the classroom, but also the collection and analysis of event data collected from the systems that are used for support and distribution (Flanagan et al., 2017; Ogata et al., 2017; Ogata et al., 2015). In addition to its advantages on students' learning, digital text readers are capable of recording interactions regarding students' reading behaviors. As the materials are read by students using the system, the action events are recorded, such as: flipping to the next or previous page, jumping to different pages,

memos, comments, bookmarks, and drawing markers to indicate parts of the learning materials that learners think are important or find difficult.

Despite the increase in use, research analyzing students' interaction with digital textbooks is still limited. Recent review study (Peña-Ayala et al., 2014) revealed that almost half of the papers in Learning Analytics (LA) and Educational Data Mining (EDM) fields are using data from Intelligent Tutoring Systems (ITS) or Learning Management Systems (LMS). Previous research into the reading behavior of students has been used in review patterns, visualizing class preparation, behavior change detection, and investigating the self-regulation of learners (Yin et al., 2015; Ogata et al., 2017; Shimada et al., 2018; Yamada et al., 2017). The analysis of reading behavior can be used to inform the revision of learning materials based on previous use, predict at-risk students that may require intervention from a teacher, and identify learning strategies that are less effective and provide scaffolding to inform and encourage more effective strategies. The digital learning material reader can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

The main motivation of this workshop is to foster research into the analysis of students' interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. This workshop builds upon previous events that have focused on student performance prediction based on reading behavior. Recently, a small dataset of e-book reading logs was used to predict the final grade score of learners in the 5th ICCE workshop on Learning Analytics (LA) & Joint Activity on predicting student performance (Flanagan, 2018), and 16 participant submissions were received. In this workshop, we offered a unique opportunity for participants to:

- Analyze **large-scale reading log data on over 1,000 students** from one university.

- Examine anonymized reading log datasets from three different universities with the possibility of cross institute analysis.

- Investigate preview, in-class, and post-class reading behaviors by analyzing the scores from quizzes/exams/final grades and lecture schedules that were provided as part of the datasets.

Aligned with LAK'19 interest in "ways in which learning analytics can be used to promote inclusion and success", in this workshop we encouraged participants to pay special attention to how different groups within the student cohort can be identified and analysis of reading logs can provide hints for how groups can be scaffolded appropriately to achieve success.

2   OBJECTIVES

While organizing the workshop, emphasis was placed on the following topics which the organizers feel due attention should be paid. Low retention and high failure rates are important problems in education (Villagrá-Arnedo et al., 2017). However, studies have shown that timely interventions for at-risk students can be effective in helping change their behaviors (Arnold et al., 2012; Tanes et al., 2011). Therefore, focusing on the early detection of at-risk students is an essential step to changing student's behavior for greater success. This broader task may be approached form the following perspectives: student reading behavior self-regulation profiles spanning the entire course,

preview/in-class/review reading patterns, student engagement analysis and behavior change detection, and visualization methods to inform and provide meaningful feedback to stakeholders.

Participants have been encouraged to contribute their programs/source code created in the workshop to an ongoing open learning analytics tool development project for inclusion as an analysis feature. This integration of research conducted in this workshop into open learning analytics infrastructure will be managed by the organizes as an ongoing effort.

3   OVERVIEW

This workshop is held in a mini-track style with a focus on presentations from participant submitted papers that analyze the data provided by the workshop. While the organizers encourage the analysis of the workshop dataset it was not exclusive, and comparative studies that also examine other sources of data were welcomed. The workshop schedule starts with a general overview of the datasets, activities and tasks, followed by presentations of papers that were submitted by participants. A total of 13 papers were accepted for publication in the companion proceedings.

A majority of contributions focused on the analysis of reading behaviors and prediction of learning outcome scores that were provided in the workshop dataset. Several different techniques were explored, such as: regression, PCA, deep learning, descriptive statistics, and stochastic block model. As with previous workshops that analyzed similar datasets, the engineering of features that accurately describe certain behaviors was discussed in detail.

Some contributions focused on the visualization of results. Minematsu et al. (2019), visualize the distributions of quiz scores and action scores to focus on active learners that achieved low quiz scores. Owatari et al. (2019), examine the reading behavior of so-called power users when compared to average learners. They found that the power user sub-group tended to flip pages earlier than other learners.

Other contributions examined topics other than the analysis of the workshop dataset. Hasnine et al. (2019), introduces a dataset of logs collected from the SCROLL informal learning tool that could have potential for the analysis and sharing of past learning experiences in foreign language learning. Flanagan et al. (2019), introduces a framework for the analysis of reading patterns from the perspective of knowledge that is contained in the learning materials being read.

ACKNOWLEDGEMENTS

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270). ACM.

Flanagan, B., & Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In *International Conference on Computers in Education (ICCE2017)* (pp. 333-338).

Flanagan, B., Chen, W., & Ogata, H. (2018). Joint Activity on Learner Performance Prediction using the BookRoll Dataset. In *International Conference on Computers in Education (ICCE2018)* (pp. 480-485).

Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19).*

Hasnine, M.N., Ogata, H., Akcapinar, G., Mouri, K., & Uosaki, N. (2019). Learning Analytics to Share and Reuse Authentic Learning Experiences in a Seamless Learning Environment. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19).*

Minematsu, T., Shimada, A., Taniguchi, R. (2019). Analytics of the relationship between quiz scores and reading behaviors in face-to-face courses. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19).*

Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M., & Kojima, K. (2017). M2B System: A Digital Learning Platform for Traditional Classrooms in University. *Practitioner Track Proceedings* (pp.155-162).

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).

Owatari, T., Shimada, A., Matsumine, T., Taniguchi, R. (2019). How Students Flip Pages during Lectures? Comparison between Power Users and Normal Users. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19).*

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.

Shimada, A., Taniguchi, Y., Okubo, F., Konomi, S., & Ogata, H. (2018). Online change detection for monitoring individual student behavior via clickstream data on E-book system. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. (pp. 446-450).

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. *Computers & Education*, *57*(4), 2414-2422.

Villagrá-Arnedo, C. J., Gallego-Durán, F. J., Llorens-Largo, F., Compañ-Rosique, P., Satorre-Cuerda, R., & Molina-Carmona, R. (2017). Improving the expressiveness of black-box models for predicting student performance. *Computers in Human Behavior*, *72*, 621-631.