Mining Sequences of Gameplay for Embedded Assessment in Collaborative Learning

Philip Buffum
North Carolina State University
Computer Science
psbuffum@ncsu.edu

Megan Frankosky
North Carolina State University
Psychology
rmhardy@ncsu.edu

Kristy Elizabeth Boyer
University of Florida
Computer & Information Science & Engineering
keboyer@ufl.edu

Eric Wiebe
North Carolina State University
STEM Education
wiebe@ncsu.edu

Bradford Mott
North Carolina State University
Computer Science
bwmott@ncsu.edu

James Lester
North Carolina State University
Computer Science
lester@ncsu.edu

ABSTRACT

This poster presents a sequence mining analysis of collaborative game-based learning for middle school computer science. Using pre-post test results, dyads were categorized into three groups based on learning gains. We then built first-order Markov models for the gameplay sequences. The models perform well for embedded assessment, classifying gameplay sequences with 95% accuracy according to whether the group learned the target concepts or not. These results lay the groundwork for accurate embedded assessment of dyads in game-based learning.

Keywords

Embedded assessment; game-based learning; collaboration; Markov models

1. INTRODUCTION

There is growing recognition of the importance of collaborative learning, in which students work together to solve problems [2, 3]. Collaboration, furthermore, can have an especially beneficial impact in game-based learning, where it has been shown to promote significant student learning gains [4] and provide significant motivational benefits [8], as well as deliver more equitable gaming experiences for diverse learners [1, 6].

Yet collaborative learning presents unique challenges to educational data mining research. While much current work in this field relies on mapping individual students' outputs, student collaboration produces learning that plays out as a joint activity, necessitating different approaches to understanding the underlying processes [7]. Recent work in educational data mining has demonstrated some success in predicting student outcomes in paired learning, as long as both students in the pair have similar initial knowledge [5].

This poster examines collaborative game-based learning in the context of the ENGAGE game-based learning environment, with which middle school students learn about computer science through an overarching narrative situated within a fictional underwater research station. In this study, students played ENGAGE in pairs at a single computer, taking turns with one set of game controls. These two students' inputs were therefore captured within a single gameplay log. The analysis presented here investigates a variation on the traditional learning question of, "Did student *S* learn the concept?" and instead asks, "Did the collaborative partnership *P* result in learning?" By building first-order Markov models on dyads' gameplay logs, we discovered that the gameplay sequences of dyads in which some learning occurred (i.e. at least one of the students learned the material) differed significantly from those in which no learning occurred, and moreover, that we can classify with very high accuracy the learning that occurred on a targeted learning objective.

2. COLLABORATIVE LEARNING TASK

This study focuses on a subset of the ENGAGE game. In ENGAGE'S Digital World level, students learn how computers process data using the binary number system. The current analysis focuses on one room in the game world, in which students integrate the two concepts of *variables* and *binary numbers*, having earlier explored both these individual concepts in isolation from one another. 124 middle school students played the game in pairs; as there is one gameplay trace for each dyad, this produced 62 gameplay traces. We administered individual pre- and post-tests to each student so that we could characterize each student's learning outcomes. The goal of the present analysis is to utilize gameplay logs to predict learning, specifically to investigate how the gameplay of those dyads who scored higher on learning assessments differs from the gameplay of those who did not score higher. Accordingly, having assigned each *individual* student a grade based on pre and post test scores, we then classified student *pairs* into one of three categories: *Learner* (19 dyads), *Prior Mastery* (23 dyads), and *Non-Learners* (20 dyads).

3. RESULTS

The modeling approach aims to identify differences in gameplay sequences between students in the *Learner*, *Prior Mastery*, and *Non-Learner* groups. We began with one of the simplest sequential models of all, first-order observable Markov models. It was expected that more sophisticated models, such as hidden Markov models or Conditional Random Fields, may be needed to characterize the gameplay sequences well; however, as this poster demonstrates, the simplest model was able to classify the gameplay sequences of *Learner*, *Prior Mastery*, and *Non-Learner* groups with high accuracy.

We built separate models for each group (*Learner, Prior Mastery, Non-Learner*) and then determined whether there were significant differences in the models for each group by comparing model fit (in terms of log-likelihood, since the probabilities themselves are very small in magnitude). We performed this pairwise comparison for all three groups, as described below:

1. For each gameplay trace sequence s_i in the Learner group:
    i. Compute $logProb(s_i \mid L_{leave-i-out})$ of observing s_i under the Learner model L (trained in a leave-one-out fashion where s_i was the left-out sequence).
    ii. Compute the log-likelihood $logProb(s_i \mid PM)$ of observing s_i under the Prior Mastery model PM trained on all Prior Mastery gameplay sequences.
    iii. Compute the log-likelihood $logProb(s_i \mid NL)$ of observing s_i under the Non-Learner model NL trained on all Non-Learner gameplay sequences.
2. Repeat the analogous process for each gameplay sequence in the Prior Mastery and Non-Learner groups.
3. For each group's sequences, test whether the set of log-likelihoods for that group under its own model is significantly higher than the log-likelihoods for that group under the other groups' models.

The models were significantly different across *Learners*, *Prior Mastery*, and *Non-Learner* groups, as shown in Figure 1, which shows the absolute values of log likelihoods for each of the three categories. In this graph, a **lower** absolute log-likelihood indicates better model fit. For each category, the graph shows three bars, the first showing the log likelihood for the given category's sequences under the *Learner* model, the second bar showing the log likelihood for the given category's sequences under the *Prior Mastery* model, and the third bar showing the log likelihood for given category's sequences under the *Non-Learner* model. We conducted a series of paired *t*-tests to determine, for each group, whether there were significant differences between the log likelihoods for its own model and those for the other two models. For the *Learner* group model, its own log likelihoods were found to be significantly better than the log likelihoods of the other two models at the $p < .01$ level. For both of the other two models, *Prior Mastery* and *Non-Learner*, their own log likelihoods were found to be significantly different than the other respective models with even greater significance, at the $p < .001$ level.

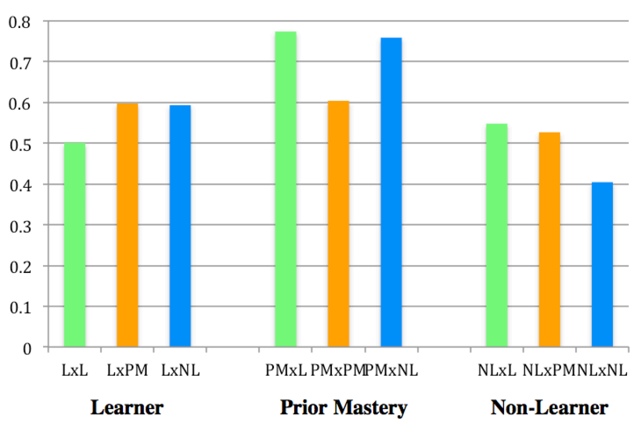

Figure 1. Absolute value of log likelihoods for each of the three categories. Lower values indicate better model fit.

Finally, we investigated the extent to which these models could classify *Learner, Prior Mastery*, and *Non-Learner* based only on the observed gameplay sequences in Room 2 and using leave-one-out cross-validation. A sequence was labeled with the group whose model produced the highest log-likelihood for that sequence (using only models that were trained with the sequence left out). Using this classifier, for the *Learner* category, 89.5% of pairs (17 out of 19) were correctly classified. For the *Prior Mastery* category, 100% of pairs (23 out of 23) were correctly classified. For the *Non-Learner* category, 95% (19 out of 20) were correctly classified. On the whole, this reflects a 95.2% accuracy in classifying whether a collaborative pair of students would be in the *Learner*, *Prior Mastery*, or *Non-Learner* group.

4. CONCLUSION

Modeling collaborative learning is an important direction for educational data mining research. We have demonstrated that sequence modeling relying on first-order Markov models can differentiate gameplay sequences of pairs where at least one partner learned from pairs who did not learn. Moreover, these models can classify those gameplay sequences with very high accuracy according to whether the dyad learned or not.

The opportunities are numerous for empirical studies into collaborative gameplay, problem solving, and dialogue. For example, the current analysis assumes that the maximal knowledge of the group is expressed through gameplay, an assumption that needs to be investigated. Additionally, a natural next step is to examine prediction power of individual learning along with the slightly more abstracted dyadic learning considered here. It is hoped that this line of investigation will move us toward highly effective support of dyadic learning.

5. REFERENCES

[1] Buffum, P.S. et al. 2016. Collaboration and Gender Equity in Game-Based Learning for Middle School Computer Science. *Computing in Science & Engineering*. 18, 2 (Mar. 2016), 18–28.

[2] Coleman, B. and Lang, M. 2012. Collaboration Across the Curriculum: A Disciplined Approach to Developing Team Skills. *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education (SIGCSE '12)* (2012), 277–282.

[3] Falkner, K. et al. 2013. Collaborative Learning and Anxiety: A phenomenographic study of collaborative learning activities. *Proceedings of the 44th ACM Technical Symposium on Computer Science Education (SIGCSE '13)* (2013), 227–232.

[4] Hickey, D.T. et al. 2009. Designing Assessments and Assessing Designs in Virtual Educational Environments. *Journal of Science Education and Technology*. 18, 2 (Feb. 2009), 187–208.

[5] Rafferty, A. et al. 2013. Estimating Student Knowledge from Paired Interaction Data. *Proceedings of the 6th International Conference on Educational Data Mining* (2013).

[6] Richard, G.T. and Hoadley, C. 2015. Learning Resilience in the Face of Bias: Online Gaming, Protective Communities and Interest-Driven Digital Learning. *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning* (2015), 451–458.

[7] Stahl, G. et al. 2006. Computer-supported collaborative learning: An historical perspective. *Cambridge handbook of the learning sciences*. (2006), 409–426.

[8] Warren, S.J. et al. 2008. A MUVE Towards PBL Writing. *Journal of Research on Technology in Education*. 41, 1 (Sep. 2008), 113–140.