

# Predicting Students Performance Based on Their Reading Behaviors

Hung Chau\*, Ang Li\* and Yu-Ru Lin

University of Pittsburgh, Pittsburgh PA 15260, USA,  
[hkc6@pitt.edu](mailto:hkc6@pitt.edu), [anl125@pitt.edu](mailto:anl125@pitt.edu), [yurulin@pitt.edu](mailto:yurulin@pitt.edu)

**Abstract.** E-learning systems can support students in the on-line classroom environment by providing different learning materials. However, recent studies find that students may misuse such systems with a variety of strategies. One particular misused strategy, gaming the system, has repeatedly been found to negatively affect the students' learning results. Unfortunately, methods to quantitatively capture such behavior are poorly developed, making it difficult to predict students learning outcomes. In this work, we tackle this problem based on a study of the 567,193 records of the 71 students' reading behaviors from two classes in the academic year 2016. We first quantify the extent to which students misused the system and then predict their class performance based on the quantified results. Our results demonstrated that such misbehavior in the E-learning system can be quantified as a probability and then further used as a significant factor to predict students class learning outcomes with high accuracy.

**Keywords:** Reading behavior, Gaming and Learning performance prediction

## 1 Introduction and Motivation

E-learning systems are one of the widely-used approaches to incorporate computer-aided teaching materials into the classroom. Many studies have demonstrated that such systems could benefit students by improving their learning efficiency [6, 8]. However, researchers have recently found that students actually choose to use such systems in various strategies, with some strategies potentially leading to poorer learning outcomes. Particularly, one strategy, gaming the system, has repeatedly been found to negatively affect students' learning results [3, 4]. As noted by Baker, gaming the system is “attempting to succeed in an educational environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly” [3]. While studies have been done to detect students gaming behavior in the E-learning system [2], it's still unclear whether this behavior could be quantified. In addition, though many studies have found that students “gaming” behavior have negative effects on their learning outcomes in the same E-learning system, it remains unclear that

---

\* First two authors have equal contribution

to what extent this quantified “gaming” measurement can be used to predict students final learning outcomes in the classroom.

Therefore, the goal of our current research is to quantify the extent to which students’ game in the E-learning system, and then investigate whether these quantified “gaming” measurements could be leveraged to predict the students learning results in the class. The specific setting for the current study is the Reading Circle system<sup>1</sup>. Students in the class are required to use Reading Circle every week for the entire semester. In addition, we use students’ final class grades to measure their learning results. Specifically, students’ final grade is the combination of their quizzes, assignments, projects and exams.

## 2 Background and Hypothesis

Many studies in educational technology have found that students choose to use E-learning systems in a variety of strategies; different strategies have been demonstrated to have distinct learning outcomes. On the one hand, good behaviors lead to better learning results. For example, the study conducted by Kinnebrew and Biswas [5] analyzed data collected from Betty’s Brain E-learning system to understand students’ learning interaction behavior. They found that students who performed better usually read the materials more carefully and systematically, having more “full” reading sections and relevant re-read actions. On the other hand, students may misuse such systems. One particular misbehavior strategy, “gaming the system”, has repeatedly been identified to have a negative effect on students’ learning outcomes. Studies have demonstrated that such behaviors are associated with students who have poor learning results [3, 7]. By controlling the students prior knowledge and general academic ability, Baker et al. [3] found that those students who frequently misused tutor software learned worse than those students who used the system properly. In a follow-up study by the same group, Baker et al. [1] built a detector to detect students “gaming” behavior under a cognitive tutor system. The model was trained on data collected from systematic classroom observations of whether a student was gaming. Using a computational gaming detector, Muldner et al. [7] also confirmed that gaming in general can be harmful to learning. Students who are less likely to game the system have better learning outcomes.

Based on these previous studies and observations, we hypothesize that “*gaming behavior in the E-learning system will have negative effect on their learning outcomes.*” We propose a method to quantitatively examine this hypothesis in this work.

## 3 Method and Data Engineering

### 3.1 Research Platform

Our current research is conducted on the Reading Circle platform. Reading Circle is a system that supports students’ learning in the on-line classroom environ-

---

<sup>1</sup> [http://adapt2.sis.pitt.edu/wiki/Open\\_Corpus\\_Personalized\\_Learning](http://adapt2.sis.pitt.edu/wiki/Open_Corpus_Personalized_Learning)

ment. It provides students with course learning materials including textbooks, research publications, etc. The system also assesses the students with multiple choice questions at the end of each section to measure students' knowledge acquisition. In addition, Reading Circle tracks and records students' entire behavior history after they log in<sup>2</sup>.

### 3.2 Method

**Identifying Gaming Behavior:** to quantify the students "gaming" behavior, we first label each reading section with "Gaming" (1) or "Normal" (0) based on their behavior. The final labeled data is about 80% inner-agreement on "gaming" behavior based on two coders who are the first two authors and were working independently. We then performed binary classification methods to calculate the probability of the "gaming" based on the features we generated (description in the next section). We test with different classification methods including Logistic Regression, KNN (with  $K = 3$ ,  $K = 5$ , and  $K = 8$ ), Naive Bayes, Decision Tree, and SVM with Linear and RBF kernel; and evaluate their performance based on *mean squared error (MSE)* metric with 10-fold cross-validation.

**Predicting Student Performance:** to predict students' final grades and class performance, we first applied regression models to analyze and select the significant predictors that influence the student's final grades. A collinearity diagnostic is also performed to avoid the potential effect of collinearity in examining the influence of predictors. Then, we used the regression model with those significant predictors to predict the student's final grades. We started with simple linear regression, and then improved the model by using nonlinear methods. Finally, the best model is able to effectively balance the variance-bias trade-off.

### 3.3 Data

The raw data was collected from two classes in the Spring 2016 semester. The system collected the total of 71 students' reading behaviors every 10 seconds for the entire semester, including 380,814 records from Human Computer Interactive (HCI) class and 186,379 from Information Retrieval (IR) class.

To effectively perform data mining on learning interaction traces, raw logs are transformed into an appropriate dataset for our two tasks. For the first task, raw data is aggregated based on each reading section, and the resulting dataset is then used for quantifying the "gaming" behavior. For the second task, the data is then aggregated by student, and the resulting dataset is used to predict the student's final performance. In both cases, each observation includes behavior features that characterize the student's interaction with the system. In the next section, we will describe the features used for two data mining tasks.

---

<sup>2</sup> Current study has IRB approval from University of Pittsburgh

### 3.4 Feature Engineer

Based on the previous studies [1, 5, 7], we extract the following features:

**Reading speed:** Number of words students read per second

**Number of words:** Number of words students read

**Number of questions:** Number of questions per reading section

**Attempt times:** Number of attempts per reading section

**First attempt correction rate:** First attempt success rate in answering questions per reading section

**Last attempt correction rate:** Last attempt success rate in answering questions per reading section

**Class:** whether this student in the HCI class or IR class

**Table 1.** Statistical Description on Behavior Features

| Variables                     | HCI: Mean | HCI: St.D. | IR: Mean | IR: St.D. |
|-------------------------------|-----------|------------|----------|-----------|
| Attempt times                 | 3.13      | 2.33       | 3.29     | 6.18      |
| First attempt correction rate | 0.38      | 0.39       | 0.48     | 0.43      |
| Last attempt correction rate  | 0.93      | 0.23       | 0.89     | 0.29      |

Two coders label each reading section as “normal” (0) or “gaming” (1) based on student’s behavior. The inner-agreement is 79.05%, including 620 gaming sections accounting for 23.90% of the dataset. Based on these features, we then calculate a new feature “**gaming**” probability which is able to identify the probability of “gaming” behavior for the student’s current section.

In our second task, we aggregate the data at the student level (based on mean and standard deviation) and predict students’ learning outcomes by incorporating both their reading behavior features and the “gaming” probability. The students’ learning outcomes are measured based on their final class performance:

**Final Grading:** Final grade of this focal student, ranging from 0% to 100%.

**Class Performance:** Separate the students final grading into binary (above average versus below average performance) based on the mean grade of the class.

**Table 1** summarizes the basic statistics of the students reading behavior features. Interestingly, students from both classes like first guessing the answer several times. The average attempt times is around 3 times with the standard deviation of 6 for IR class. Both classes have correction rate for first attempt (38%, 48%) much lower than it for last attempts (93%, 89%).

## 4 Results and Discussion

### 4.1 Gaming Probability

In this section, we report the performance of classification methods and choose the best model to provide the gaming probability for predicting student performance. As can be seen from **Table 2**, the Decision Tree model outperforms all the others, with the lowest MSE (0.0196). Hence, we choose the Decision Tree model to build our gaming probability provider.

**Table 2.** Model performance on detecting gaming behaviors

| Methods                | MSE           |
|------------------------|---------------|
| Logistic Regression    | 0.0464        |
| K Nearest Neighbor K=3 | 0.0655        |
| K Nearest Neighbor K=5 | 0.0661        |
| K Nearest Neighbor K=8 | 0.0711        |
| Naive Bayes            | 0.059         |
| <b>Decision Tree</b>   | <b>0.0196</b> |
| SVM Linear             | 0.0416        |
| SVM RBF                | 0.0386        |

**Table 3.** Students Learning Outcomes

| Variables                      | Coef.             | P-value         |
|--------------------------------|-------------------|-----------------|
| Ave. Reading Speed             | -1.975e-04        | 0.157           |
| St.D. Reading Speed            | 1.531e-04         | 0.066           |
| Ave. first attempt cor. rate   | -7.314e-02        | 0.058           |
| St.D. first attempt cor. rate  | -7.483e-03        | 0.918           |
| Ave. last attempt cor. rate    | 1.369e-01         | <0.05           |
| St.D. last attempt cor. rate   | 1.196e-01         | <0.05           |
| <b>Ave. gaming probability</b> | <b>-1.621e-01</b> | <b>&lt;0.05</b> |
| St.D. gaming probability       | 1.169e-01         | 0.127           |
| Class HCI                      | -2.515e-2         | 0.068           |

## 4.2 Effect of Students “Gaming” Behavior

**Table 3** presents the results in predicting student’s final class grade. First, the independent variables are not correlated, with a *Variance Inflation Factor* (VIF) value less than 5. The regression results demonstrate that the students reading behavior has significant effects in their final grades. Specifically, we find that students who have higher probability of “gaming” are less likely to achieve better grades in their class ( $Coef. = -1.62e - 01, p < 0.05, 95\% CI [-3.16e - 01, -8.39e - 03]$ ). These results support our hypothesis that the presence of the “gaming” behavior in the E-learning system can have negative influence on students’ learning results.

In addition, we find that students who have higher **last** attempt correction rate when answering the question ( $Coef. = 1.369e - 01, p < 0.05, 95\% CI [2.243e - 02, 2.513e - 01]$ ) have better learning results. Whereas, surprisingly, we find that higher **first** attempt correction rate in question and answering process ( $Coef. = -7.314e - 02, p < 0.05, 95\% CI [-1.490e - 01, 2.793e - 03]$ ), less likely the student has better grade.

## 4.3 Students’ Performance Prediction

We create regression models and classification models on those significant factors to predict students’ learning results. By first using only the linear regression with leave-one-out cross validation, the model could predict students’ final grades with root mean square error (RMSE) as 0.048. The diagnostics suggest that though the model met the normality, independence and constant variances assumptions, residuals vs. fitted graph showed a curved relationship. Therefore, we improve the model by adding quadratic terms with regularization methods to avoid the over-fitting. Eventually, we are able to reduce the RMSE to 0.046.

We then use classification models to predict students’ final performance. Our results demonstrate that Support Vector Machine (SVM) using Radial kernel performed the best with accuracy as 80% and AUC as 83%.

## 5 Conclusion and Discussion

In this work, we first quantified the extend to which students misbehaved in the E-learning system. Leveraging the classification models, we were able to

calculate the probability that students “game” the system for every reading section. The final model’s performance was very impressive with the MSE around 0.0196. Based on the quantified misbehavior measurements, we then tested our hypothesis that the presence of students’ “gaming” behavior leads to poorer learning results. Leveraged on the behavior features that we identified, we then demonstrated that students’ final grades and final performance can be predicted with over 80% accuracy rate.

In addition to understanding and quantifying the students’ “gaming” behavior, our research can also inform the current E-learning system with some design implementations, such as providing students with personalized guidance, reminders and suggestions based on their studying progress. For example, in case they attempt to “game” the system, we should send him/her reminders; or if we know the student will perform poorly in the class, we should provide suggestions. One limitation of our current work is that we only tested out this method in the data collected from academic year 2016. As part of the future work, we plan to collect more data from more class cohorts, and then test our models to see whether they can be generalized across other classes.

## References

1. Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why students engage in gaming the system behavior in interactive learning environments. *Journal of Interactive Learning Research* 19(2), 185–224 (2008)
2. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting student misuse of intelligent tutoring systems. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) *Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science*. vol. 3220, pp. 531–540. Springer, Heidelberg (2004)
3. Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z.: Off-task behavior in the cognitive tutor classroom: When students “game the system”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 383–390. CHI '04, ACM, New York, USA (2004)
4. Baker, R.S., Roll, I., Corbett, A.T., Koedinger, K.R.: Do performance goals lead students to game the system? In: *Proceedings of the 2005 Conference on Artificial Intelligence in Education: Supporting Learning Through Intelligent and Socially Informed Technology*. pp. 57–64. IOS Press, Amsterdam, The Netherlands (2005)
5. Kinnebrew, J.S., Biswas, G.: Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In: *Proceedings of EDM 2012*. pp. 57–64. Chania, Greece (2012)
6. Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A.: Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)* 8(1), 30–43 (1997)
7. Muldner, K., Bursleson, W., Van de Sande, B., VanLehn, K.: An analysis of students’ gaming behaviors in an intelligent tutoring system: predictors and impacts. *User Modeling and User-Adapted Interaction* 21(1), 99–135 (2011)
8. Suchman, L.A.: *Plans and Situated Actions: The Problem of Human-machine Communication*. Cambridge University Press, Cambridge, UK, 2 edn. (1987)