Predicting Student’s Performance Based on Their Reading Behaviors

Hung Chau*, Ang Li*, Yu-Ru Lin
School of Information Science, University of Pittsburgh, Pittsburgh, PA

Introduction and Motivation

- E-learning systems are one of the widely-used approaches to incorporate computer-aided teaching materials into the classroom
- However, researchers recently find that students actually choose to use such systems in variety strategies, with some strategies potentially leading to poorer learning outcomes.
- “Gaming” the system indicates that students attempt 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.

Background and Hypothesis

- Previous Study
  - Students choose to use E-learning systems in a variety of strategies; different strategies have been demonstrated to have distinct learning outcomes.
  - Students who performed better usually read the materials more carefully and systematically.
  - Students who frequently misused tutor software learned worse than those students who used the system properly.
- Hypothesis
  - “Gaming behavior in the E-learning system will have negative effect on students’ learning outcomes.”
- We propose a method to quantitatively examine this hypothesis in this work.

Research Platform

Reading Circle Platform

Method, Data, and Features

- Method
  - Identifying Gaming Behavior:
    - Using binary classification methods to calculate the probability of “gaming.”
    - Evaluating their performance based on mean squared error (MSE) metric with 10-fold cross-validation.
- Data Collection
  - The 71 students’ reading behaviors were recorded every 10 seconds for the entire semester.
  - Total of 380,814 records from Human Computer Interaction (HCI) class and 186,379 Information Retrieval (IR) class.
- Feature Engineer
  - Feature Set

Method, Data, and Features (Continued)

- Students from IR classes read faster than HCI class.
- Anova test showed two means are significantly different.

Results - “Gaming” Behavior

- Students who have higher last attempt correction rate when answering the question (Coef. = -1.62e 01, p<0.05, 95% CI [-3.16e 01, -8.39e 03]).
- Students who frequently misused tutor software learned worse than those students who used the system properly.
- Using binary classification methods to calculate the probability of “gaming.”
- Evaluating their performance based on mean squared error (MSE) metric with 10-fold cross-validation.
- Predicting Student Performance:
  - Class grade prediction: regression methods.
  - Class performance: classification methods.
- Data Collection
  - The 71 students’ reading behaviors were recorded every 10 seconds for the entire semester.
  - Total of 380,814 records from Human Computer Interaction (HCI) class and 186,379 Information Retrieval (IR) class.
- Feature Engineer
  - Feature Set

Results - Gaming Probability

- Decision Tree model outperforms all the others, with lowest MSE (0.0196).
- We choose Decision Tree model to build our gaming probability provider.

Results - Students’ Performance Prediction

- Class Grades Prediction: regression on significant features and quadratic terms with regularization methods. Outperforms all other methods, with lowest MSE (0.046).
- Students performance Prediction (0/1): Support Vector Machine (SVM) using Radial kernel in predicting students performance outperforms the other methods with accuracy as 80% and AUC as 83%.

Conclusion

- Leverage 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.
- Leveraged on the classification models, we were able to calculate the probability that students “gamed” the E-learning system.
- Based on “gaming” probability, we then tested our hypothesis that the presence of students “gaming” behavior leads to poorer learning results.
- Leverage 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.