Is Gaming the System State-or-Trait? Educational Data Mining Through the Multi-Contextual Application of a Validated Behavioral Model

Ryan S.J.d. Baker

Learning Sciences Research Institute
University of Nottingham
Nottingham, UK
ryan@educationaldatamining.org

Abstract. In this paper we discuss the use of a validated behavioral model across multiple contexts. We show that such a model can be used to distinguish between classes of explanations for why that behavior occurs. Specifically, we compare between state and trait explanations for why students game. We use the behavior model to predict each student's gaming frequency in a set of 35 tutor lessons, and then use linear models and the Bayesian Information Criterion to determine which class of explanations predicts gaming behavior more successfully.

1 Introduction

In recent years, it has been repeatedly documented that students choose to interact with interactive learning environments in an impressive variety of ways. Some students avoid asking for help at all costs [1], some students game the system [4,8,12], and some students even work thoughtfully and carefully in order to learn the material [1,3,7]. In recent years, a variety of models have been developed which can detect many of these behaviors [1,3,4,7], and some of these models have been incorporated into learning environments which use the models' assessments to respond to differences in student behavior [5,13].

However, there is a question that is fundamental to developing systems that can respond to differences in student behavior: why. Why do students choose to use learning environments differently from each other? And within this, why does a specific student choose to engage in a specific behavior? For example: Why did student 73 choose to game the system?

Broadly, there are two types of potential explanations for why a specific person engages in a specific behavior: **state** explanations, and **trait** explanations. State explanations suggest that some aspect of the student's current state or situation guide a student to engage in that behavior. Trait explanations, by contrast, suggest that specific traits that a student has – such as personality characteristics or preferred meta-cognitive strategies – guide a student to engage in that behavior. Trait explanations can include both fairly fixed traits (such as personality characteristics or learning disabilities) and more fluid traits (such as attitudes or preferred meta-cognitive strategies).

Several studies in recent years have attempted to correlate both state and trait explanations with student behavior in interactive learning environments [cf. 3,7,12], combining student responses on questionnaires with some indicator of their behavior within an interactive learning environment. These studies have found that a wide variety of different factors, both state and trait, are associated with specific student behaviors: however, the correlations have generally been low. For example, across Baker et al [7] and Walonoski and Heffernan [12], seven different explanations (4 state, 3 trait) were found to be statistically significantly associated with gaming the system, but none with an r^2 greater than 0.07.

An account which only achieves an r^2 of 0.07 can not be considered a primary account for why the behavior occurs. Hence, current approaches do not appear to have made large headway on resolving fundamental questions about why students choose specific behaviors in learning environments.

In this paper, we will argue in favor of a different method for analyzing why students engage in a specific behavior: educational data mining, through the broad, multi-context application of a validated model of student behavior. We will use an existing model of a category of student behavior to predict that behavior's incidence among a substantial number of students. More importantly, we will use that model to predict student behavior across a wide variety of contexts (note that this depends upon a model which has been validated across the full variety of contexts). We will show that this method is effective for distinguishing between the relative impact of state and trait explanations for student behavior, and discuss how this method can be expanded to analyze a large set of explanations quickly and efficiently.

In this paper, we focus specifically on a category of behavior known as gaming the system. Gaming the system is defined as attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material [4]. Gaming has been found to split into two distinct categories of behavior, one of which is associated with significantly poorer learning [4]. As already mentioned, gaming the system has been found to be statistically significantly associated with a variety of state and trait explanations [cf. 7,12] but those explanations have in all cases achieved low r^2.

2 Data

In order to analyze whether state explanations or trait explanations are better predictors of whether a student will game the system, we obtained data for 240 students' use of a Cognitive Tutor curriculum [2] for middle school mathematics, during an entire school year (August 2001-May 2002). All of the students were enrolled in mathematics classes in one middle school in the Pittsburgh suburbs which used Cognitive Tutors two days a week as part of their regular mathematics curriculum, year round. None of the classes were composed predominantly of gifted or special needs students. The students were in the 6^{th}, 7^{th}, and 8^{th} grades (approximately 10-13 years old).

Each of these students worked through a subset of 35 different lessons within their Cognitive Tutor curriculum, covering a diverse selection of material from the middle school mathematics curriculum. Middle school mathematics, in the United States, generally consists of a diverse collection of topics, and these students' work was representative of that diversity, including lessons on combinatorics, decimals, diagrams, 3D geometry, fraction division, function generation and solving, graph interpretation, probability, and proportional reasoning. On average, each student completed 13.7 tutor lessons (SD = 4.9), for a total of 3292 student/lesson pairs.

In the analyses presented here, we will analyze whether state explanations or trait explanations are better at predicting whether a student will game the system in a fashion associated with poorer learning [cf. 4]. To determine how often each student gamed the system, in each lesson, we applied a detector of a sub-category of gaming behavior associated with poorer learning [cf. 4,6] to a data set composed of each action by each student, in each of the 35 lessons. The data set was composed of approximately 804,000 actions in the tutor, which equaled 182.9 MB of distilled data in a flat database, or 407 MB of log files prior to distillation. The gaming detector is structurally a Latent Response Model [8]. It assesses gaming by first making predictions about whether each individual action is an instance of gaming, and then aggregates these predictions in order to make coarser grain-size predictions about how often each student games the system in each lesson. The detector was trained using data from five tutor lessons (300 students, using the tutor from 2003-2005) drawn from the same middle school mathematics curriculum as the lessons used in the analysis reported in this paper.

Since the detector was trained using data from four tutor lessons, and is being applied to data from thirty-five lessons, it is reasonable to ask whether the detector will produce reliable estimates of gaming frequency in the lessons it was not trained on. In this case, we can have reasonably high confidence, because the detector has been validated to transfer to new tutor lessons it was not trained on, within this specific tutor curriculum for middle school tutor mathematics. In [6], the gaming detector was trained on three lessons and tested it on a fourth lesson, in four different combinations. All four lessons were drawn from the same middle school tutor curriculum as the thirty-five lessons are drawn from. The detector transferred to lessons it was not trained on with only mild and non-statistically significant degradation in performance. Since all lessons used in the analysis here are drawn from that same curriculum, we have reason to believe that the detector, in general, should be reliable for the lessons studied in this analysis.

Hence, the detector gives us a prediction for gaming frequency for 3292 student/lesson pairs, which we can use to study whether gaming frequency is better predicted through state explanations or trait explanations.

3 Analysis and Results

We can determine the relative effectiveness of state and trait explanations, by setting up regression models that attempt to predict each student/lesson gaming frequency using a function on either the student, or the lesson. In other words, we treat both student and lesson as nominal variables, assign each student and/or lesson a value, and attempt to predict the gaming frequency associated with each student/lesson pair. Student is a good proxy for all trait explanations put together, because the sum total of each student's traits should be expressable as one value for that student. Similarly, lesson is a good proxy for all state explanations put together, because the sum total of a number of contextual factors should differ lesson-by-lesson and thus should be expressable as a single value for each lesson. (To explain this another way, imagine a model with 8 trait variables and 8 state variables; each student will have a weighted sum value for those 8 trait variables, and each lesson will have a weighted sum value for those 8 state variables).

Hence, we can attempt to predict gaming behavior with trait explanations by assigning a term to each student, i.e.

$$\texttt{Gaming Frequency = Student + } \alpha_0$$

The resulting model has 240 parameters (240 students). The model achieves a moderately low r^2 of 0.16, with a Bayesian Information Criterion (BiC) value of 1382. BiC values greater than zero mean that a model is over-fit [10], which suggests that despite the fact that the model's r^2 is moderately above zero, the model is in fact somewhat worse than what would be expected, by chance, from a model with 240 parameters.

We can attempt to predict gaming behavior with state explanations by assigning a term to each lesson, i.e.

$$\texttt{Gaming Frequency = Lesson + } \alpha_0$$

The resulting model has 35 parameters (35 lessons). The model achieves a considerably better r^2 of 0.55, with a Bayesian Information Criterion (BiC) value of -2370. BiC values lower than zero mean that a model predicts the data better than could be expected by chance. The distribution of gaming frequencies, lesson by lesson, is shown in Figure 1.

In addition, the large difference between the BiC values indicate that the trait model is a significantly better of gaming frequency than the state model. A difference of 10 is considered to be evidence equivalent to a p value of 0.01 [10]; these two models' BiC values differ by 3,752.

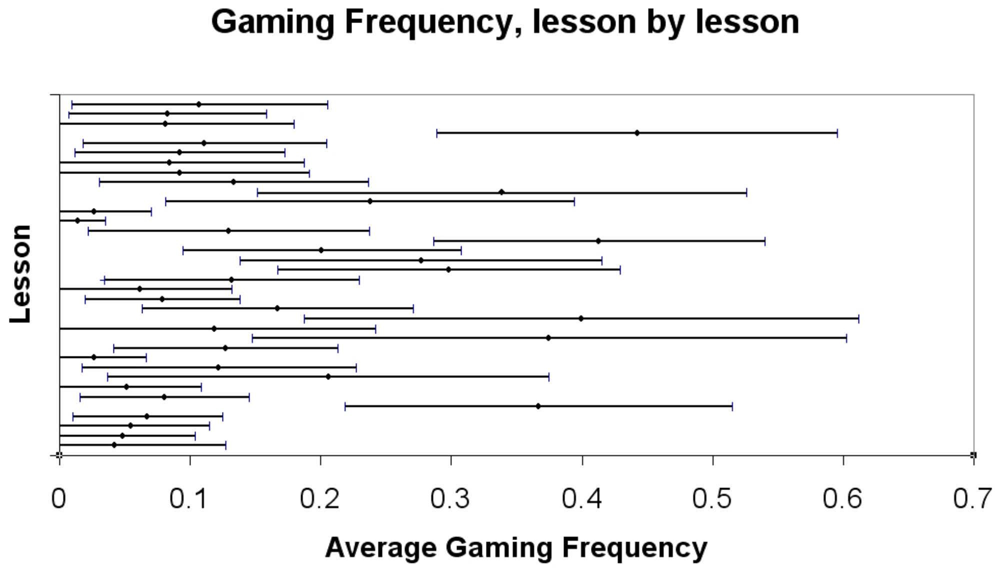

Fig. 1. Gaming frequency across lessons. Standard deviation bars used instead of standard error bars, in order to show distribution of data rather than statistical significance of difference between groups.

4. Discussion and Conclusions

The models presented here suggest that gaming the system can be generally better understood through state explanations than trait explanations. This suggests that, in order to understand why students game the system, it will be more fruitful for future work to investigate state explanations, rather than trait explanations.

In addition, the relationship found between state explanations and gaming behavior (r^2 of 0.55) is much stronger than any of the relationships found through more traditional methods of research (r^2 under 0.07). This suggests that the analytical method used here may be more powerful than previous methods used. Analytical methods that dig into the specific contexts, within lessons, which students game with particular frequency may be even more powerful for explaining gaming behavior.

The major difference between the analytical method used here, and prior research, is the number of lesson contexts studied. To our knowledge, previous studies of why students engage in specific behaviors in interactive learning environments either involved only a single lesson/ curricular sub-section [cf. 3,7] or had data from multiple lessons/curricular sub-sections, but used an overall measure of the behavior, which did not make distinctions at the lesson-by-lesson level [cf. 12]. By contrast, the study reported here involved 35 different lessons.

Using data from multiple tutor lessons gives substantial leverage for assessing both state and trait explanations. For assessing state explanations, traditional methods have involved either asking questions that attempt to assess a state's frequency or existence across the entire use of a system [cf. 3,12], or periodic assessments of a student's state across a limited amount of time [cf. 11]. Assessing student behavior across a wide variety of states, which is made possible through applying a validated model to many tutor lessons or curricular sub-sections, will inherently have higher power than such traditional approaches.

For assessing trait explanations, dividing data by lessons also gives additional statistical power. Any effective trait explanation should be an effective predictor across multiple contexts. Treating each individual student as a separate predictor of gaming is the strongest possible trait-based explanation of why students game. The fact that this predictor only achieved an r^2 of 0.16 is quite strong evidence that trait explanations will not provide the most important explanations for why students game the system.

The analysis discussed here has taken a very high-level view of state explanations and trait explanations. An important area of future work will be to apply these methods to more precise questions, involving individual elements of a student's state. The analysis presented here suggests that states – in general – are an important predictor of why students game. The large, broad, and most importantly, labeled data set that was necessary to conduct the analysis given here will in the future make it possible to conduct very sensitive comparisons of how different aspects of a student's state affects their likelihood of engaging in gaming behaviors.

In the next few years, we believe that the combination of large, broad data sets with models validated across multiple contexts will create a situation where relatively simple techniques for data exploration (such as regression and criterion-based model selection) can answer fundamental questions about why students choose to use learning environments in the ways they do.

Acknowledgements. I would like to thank Angela Wagner and Albert Corbett for enabling access to the data used in the analyses here, and Ido Roll, Albert Corbett, and Ken Koedinger for their role in the original development of the detector of gaming behavior used in these analyses. The work presented in this paper was funded by a research fellowship from the Learning Sciences Research Institute, at the University of Nottingham.

References

1. Aleven, V., McLaren,B.M., Roll, I., and Koedinger, K.R. Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS 2004),* 227-239.
2. Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences,* 4 (2), 167-207.
3. Arroyo, I., Woolf, B. (2005) Inferring learning and attitudes from a Bayesian Network of log file data. *Proceedings of the 12th International Conference on Artificial Intelligence in Education,* 33-40.
4. Baker, R.S., Corbett, A.T., and Koedinger, K.R. (2004) Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of the 7th International Conference on Intelligent Tutoring Systems,* 531-540.
5. Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J. (2006) Adapting to When Students Game an Intelligent Tutoring System. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 392-401.
6. Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Roll, I. (2006) Generalizing Detection of Gaming the System Across a Tutoring Curriculum. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems,* 402-411.
7. Baker, R.S., Roll, I., Corbett, A.T., Koedinger, K.R. (2005) Do Performance Goals Lead Students to Game the System? *Proceedings of the 12th International Conference on Artificial Intelligence in Education,* 57-64.
8. Beck, J.E. (2005) Engagement Tracing: using response times to model student disengagement. *Proceedings of the 12th International Conference on Artificial Intelligence in Education*, 88-95.
9. Maris, E. Psychometric Latent Response Models. *Psychometrika*
10. Raftery, A.E. (1995) Bayesian Model Selection. Sociological Methodology, 111-196.
11. Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S, Sugay, J.O., Tep, S., Viehland, N.J.B. (submitted) Affect and Usage Choices in Simulation Problem Solving Environments.
12. Walonoski, J.A., Heffernan, N.T. (2006a) Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems,* 382-391.
13. Walonoski, J.A., Heffernan, N.T. (2006b) Prevention of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the 8th International Conference on Intelligent Tutoring Systems,* Springer-Verlag: Berlin. pp. 722-724.