

# Variable Construction and Causal Discovery for Cognitive Tutor Log Data: Initial Results

Stephen E. Fancsali

Department of Philosophy  
Carnegie Mellon University  
& Apollo Group, Inc.

sfancsali@gmail.com

## ABSTRACT

We present a method to simultaneously search for student-level variables constructed from Cognitive Tutor log data and graphical causal models. We seek causal explanations of behavior in Cognitive Tutors, including “gaming the system” and off-task behavior, selecting variables by their contribution to causal structure and strength learning.

## Keywords

causality, intelligent tutors, discovery with models

## 1. INTRODUCTION<sup>1</sup>

Researchers constructing student-level statistical and causal models from “raw” log data of courseware must construct variables to represent student-level, aggregate features of interest. We propose search for constructed variables, with a focus on “gaming the system” and off-task behavior in Cognitive Tutors. Variables are assessed by their support of inferences about causation and causal strength in graphical causal models [12].

“Feature engineering” has been explored for predictive models of educational data (e.g., [1], [4]). Usually targets (not always, cf. [6]) have been fine-grained outcomes (e.g., success at next tutor interaction) and not student-level outcomes (e.g., exam scores).

Graphical causal models have also been used on educational data ([10], [11], etc.) with student-level features. This work develops an approach [9] that combines data-driven variable construction and algorithmic causal discovery to model student behavior.

## 2. DATA + MOTIVATION

Data are from interactions of 102 non-traditional, adult learners with the Carnegie Learning Algebra Cognitive Tutor in an (online or on-campus) algebra course at the University of Phoenix, specifically data from the last module of the course. Target learning outcomes are students’ course final exam scores.

Learners “game the system” by taking advantage of intelligent tutor properties to get through course material without genuinely learning [2]. Off-task learners disengage from the tutor and behave in ways unrelated to learning tasks [7]. Both types of behavior have been associated with decreased learning [3].

Research on gaming describes it as “harmful” in a non-causal way, denoting mere association with negative outcomes. Methods for inferring causal relationships from observational data may

help provide evidence for (or against) a causal relationship between gaming behavior and learning.

We deploy software “detectors” of gaming [5] and off-task behavior [7] that use a variety of “engineered”/“distilled” features [4] to determine whether a transaction corresponds to gaming or off-task behavior. We treat their output as “fine-grained” observations of behavior and seek variables to represent aggregate behavior over “fine-grained” observations.

Recent work [8] considers whether influences of gaming the system and off-task behavior on learning are immediate, aggregate or both. They construct variables over lessons/units of interest and report that gaming is weakly associated with aggregate poorer learning and that off-task behavior is strongly associated with aggregate poorer learning.

## 3. METHOD + RESULTS

Our data range over several units (with corresponding sections) of algebra material and 32 skills. Student behavior in particular units, sections, or skills is possibly more important for learning than behavior over the entire module. We search over variables constructed at these levels of aggregation. Models in [8] consider variables as *counts* of gamed/off-task steps. As they suggest, other functions might manifest important behaviors. Our strategy is to search over constructed variables to find those that support causal inferences using algorithmic search for causal models.

Graphical causal models are frequently directed acyclic graphs (DAGs) with associated probability distributions (Bayesian networks). DAG nodes represent variables; edges represent causal relationships. Two assumptions<sup>2</sup> link causal structure represented by a DAG to independencies entailed by a DAG: the Causal Markov Condition and Causal Faithfulness Condition [12].

Algorithms like FCI<sup>3</sup> [12] learn the equivalence class of graphs compatible with conditional independence relations among measured variables, assuming there may be unmeasured common causes of measured variables. FCI returns a partial ancestral graph (PAG), representing the set of causal graphs compatible with conditional independence relations among measured variables. Edge interpretations in a PAG are  $X \circ \rightarrow Y$ : Either  $X$  causes  $Y$ , or  $X$  and  $Y$  share a latent common cause (or both);  $X \circ - \circ Y$ : (1)  $X$  causes  $Y$ , (2)  $Y$  causes  $X$ , (3)  $X$  and  $Y$  share a latent common cause, or (4) either (1) & (3) or (2) & (3);  $X \leftrightarrow Y$ : There is a latent common cause of  $X$  and  $Y$ ;  $X \rightarrow Y$ :  $X$  causes  $Y$ .

To judge which variables support causal inferences with uncertain causal structure, we iterate<sup>4</sup> over DAGs consistent with a PAG

<sup>1</sup> The author gratefully acknowledges the Apollo Group, Inc., for funding as well as Ryan Baker, David Danks, Sujith Gowda, Tristan Nixon, Joseph Ramsey, Steve Ritter, Partha Saha, and Richard Scheines for assistance and advice.

<sup>2</sup> We omit illustrations of the framework and assumptions.

<sup>3</sup> The Tetrad Project <<http://www.phil.cmu.edu/projects/tetrad/>>

<sup>4</sup> Our heuristic is not ideal; a better method is work in progress.

over sets of constructed variables and calculate “average causal predictability” achieved. For each compatible DAG, we specify a linear regression model for *final\_exam* score with its direct causes as predictors. If each DAG is equi-probable, we maximize average  $R^2$  value over these models.

Fig. 1 is a baseline causal model (PAG) for variables from [8]. The negative association of *number\_steps\_gamed* and *final\_exam* is likely induced by a causal relationship. Average causal  $R^2$  (“causal predictability”) for this PAG is .5028.



Fig. 1. Baseline PAG; +/- indicate association

We search over aggregate variables constructed from characteristics tracked by Cognitive Tutors or calculated by “detectors” – counts (and when relevant, proportions) of transactions: overall, correct, wrong, help/hint request, “known bugs” (misconceptions), gamed, off-task; transaction time taken; average gaming & off-task (numerical) estimate; counts (and relevant proportions) of: steps, gamed steps, and off-task steps. We consider aggregating over the entire module and sections, units, and skills within the module. For each level, we consider functions of step-level characteristics (and their natural logarithm) to determine constructed variables: sum, average, variance, max, and min. The schema for constructed variable names is: *LEVEL\_level-name\_function(characteristic)*.

Applying functions at different aggregation levels, we “explode” a set of a few hundred variables and “prune” by removing uninformative and redundant variables; for highly correlated pairs, we remove the variable with lower target correlation. From 20 variables with highest correlation to the target, we randomly select sets of (9) variables, apply FCI, and seek the set that maximizes average causal  $R^2$  afforded by the PAG. More work is required to determine the best sizes for these variable sets.

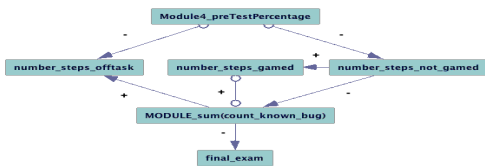


Fig. 2. Augmented baseline PAG; avg. causal  $R^2 = .5816$

Full search shows the importance of the count of misconceptions in the module: *MODULE\_sum(count\_known\_bugs)*. Augmenting our baseline variables with this, we apply FCI (PAG, Fig. 2). Our baseline model suggests the  $o \rightarrow o$  edge between *number\_steps\_gamed* and “misconceptions” can be oriented as  $\rightarrow$ ; the latter is a more proximate cause of learning.

For our full search, our training set consists of variables computed over 80% of steps randomly sampled over all students. Test set variables are computed over the remaining steps. The PAG (Fig. 3) that maximizes causal predictability for the training set has average causal  $R^2 = .614$  (test set avg. causal  $R^2 = .5137$ ).

We establish that at least one variable mediates a (likely) causal link between “gaming” and learning. Future work will extend and generalize this approach and refine the search space.

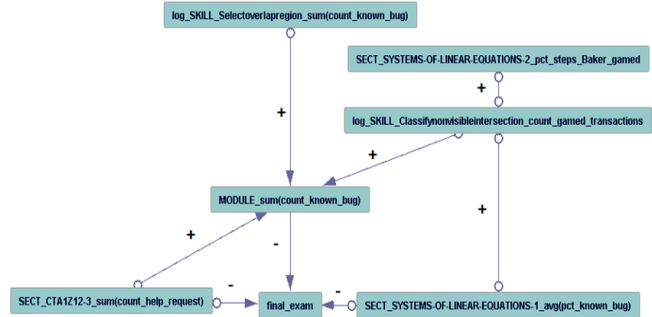


Fig. 3. PAG for full variable search

## 4. REFERENCES

- [1] Arnold, A., Beck, J., Scheines, R. Feature Discovery in the Context of Educational Data Mining: An Inductive Approach. Proceedings of the AAAI 2006 Workshop on Educational Data Mining, (2006), 7-13.
- [2] 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. Adapting to When Students Game an Intelligent Tutoring System. Proceedings of the 8<sup>th</sup> International Conference on Intelligent Tutoring Systems, (2006), 392-401.
- [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.” Proceedings of ACM CHI 2004, 383-390.
- [4] Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. Developing a Generalizable Detector of When Students Game the System. User Modeling and User-Adapted Interaction, 18 (2008), 287-314.
- [5] Baker, R.S.J.d., de Carvalho, A. M. J. A. Labeling Student Behavior Faster and More Precisely with Text Replays. Proceedings of the 1<sup>st</sup> International Conference on Educational Data Mining. (2008), 38-47.
- [6] Baker, R.S.J.d., Gowda, S.M., Corbett, A.T. Automatically Detecting a Student’s Preparation for Future Learning: Help Use is Key. Proceedings of the 4<sup>th</sup> International Conference on Educational Data Mining, (2011), 179-188.
- [7] Baker, R.S.J.d. Modeling and Understanding Students’ Off-Task Behavior in Intelligent Tutoring Systems. Proceedings of the 2007 Conference on Human Factors in Computing Systems, (2007), 1059-1068.
- [8] Cocea, M., Hershkovitz, A., Baker, R.S.J.d. The Impact of Off-Task and Gaming Behavior on Learning: Immediate or Aggregate? Proceedings of the 14<sup>th</sup> International Conference on Artificial Intelligence in Education, (2009), 507-514.
- [9] Fancsal, S.E. Variable Construction and Causal Modeling of Online Education Messaging Data: Initial Results. Proceedings of the 4<sup>th</sup> International Conference on Educational Data Mining, (2011), 331-332.
- [10] Rai, D., Beck, J.E. Exploring User Data from a Game-Like Math Tutor: A Case Study in Causal Modeling. Proceedings of the 4<sup>th</sup> International Conference on Educational Data Mining, (2011), 307-311.
- [11] Scheines, R., Leinhardt G., Smith, J., Cho, K. Replacing Lecture with Web-Based Course Materials. Journal of Educational Computing Research, 32 (2005), 1-26.
- [12] Spirtes, P., Glymour C., Scheines, R. Causation, Prediction, and Search. 2<sup>nd</sup> Edition. MIT, Cambridge MA, 2000.