Towards Building an Automated Detector of Engaged and Disengaged Behavior in Game-Based Assessments

Stephenson, S., Baker, R. and Corrigan, S.

Abstract

We present current work applying feature engineering processes and educational data mining techniques to create an automated detector of engaged behaviors in a game based assessment. The resulting detector for engaged behaviors is capable of identifying patterns of gameplay that provide evidence that players are engaged in meeting the objectives of one or more of the game's missions. It is expected that the approach to conceptualizing engaged behavior is generalizable to other GBAs.

Research Rationale & Significance

Game based assessments (GBAs) are digital games designed to support student learning and provide formative assessment information about what students know and can do (Mislevy, et al. 2014). However, development of GBAs represents substantial investment. Their value is fully realized as assessments and instructional tools only where students engage with them in ways intended by their designers. As a result, there is a pressing need among teams building high quality GBAs to detect when students are not using them as designed. The effort presented here applied feature engineering processes and educational data mining techniques in order to create an automated detector of disengaged behavior within a recently designed GBA called Pollution City. Developed by GlassLab, Pollution City repurposes the code base of Maxis' recently released SimCity. Successful completion of the game's missions requires players to solve a series of optimization problems that require complex problem solving (Shute, 2011; Brown, 2011) - simultaneous attention to and manipulation of multiple independent variables.

Research Questions and Objectives

We present work on a data-mined automated detector of engaged behavior within the context of a revised version of SimCity. Given the game's focus on teaching and assessing complex problem solving, we are interested in identifying and modeling the students' engaged behaviors as predictors of their game play success. In general, we focus on identifying which features of engagement are predictive of student success within the game and do they differ when taking game replay (multiple attempts) into account. Efforts to detect and monitor incidences of engaged behavior within Jackson City raises the following questions:

1. What are the key features of engagement related to game success among all player attempts?
2. Among students who played at least twice, are the features of engagement the same after each game play?

The first question identifies the engaged behaviors related to all students who play Jackson City any number of times. Some may have only played once while others have made up to 17 attempts! Given that the number of attempts vary so wildly, question two becomes very interesting to address. We investigate this question by first selecting the analysis sample based on students who have played more than once. We then compare the features identified from their first attempt to the second attempt. Operating under the assumption that more game play attempts can result in different levels of engagement, we expect to identify a different set of features for each attempt.

Data

The main source of data comes from a log of all student actions from their first attempt playing through the final game mission, "Jackson City". Over a span of 15 minutes, students can bulldoze buildings, place new power structures (wind, solar, or coal generated), build new roads to expand their city, and zone and dezone residential, commercial, and industrial areas in order to achieve their goals. They can also monitor the effects of their actions on pollution and jobs with the on-screen thermometers. Given a diverse set of game-play actions, we are able to build a detector to understand the relationship between cognitive engagement and success within the game. Success in the game is measured by student's ability to lower pollution levels and increase job values. To accomplish this goal, students need to suppress pollution values to at least 50 units while increasing jobs to at least 2600 for their citizens.

Method

We studied which engaged behaviors were associated with student success, using a combination of feature engineering (Sao Pedro et al., 2012) and prediction modeling. In brief, feature engineering is the process of identifying specific actions or features within the game that are likely to provide evidence about players' engagement. This form of knowledge engineering is applied within the framework of predictive data mining where we employ regression algorithms to identify the final set of engagement features that predict success in the game. Regression is a statistical measure that attempts to determine the strength of the relationship between the game success dependent variable (JCSI) and the series of independent variables (known as attributes). The outcome measure of the model includes a composite variable intended to capture the distribution of success within Jackson City (JCSI). The composite is based on a multi-step algorithm that puts scores for jobs and pollution on the same scale. It bounds all values between 0 and 1 [0,1] so that the “best” score is 1 and the “worst” score is 0.

Features of Engagement

Detector development and modeling cognitive engagement within the game environment. We engineered a set of features based on a combination of previous research on modeling engagement and theory. Our initial set of features included over 80 game-play actions as potential sources of evidence about players’ engagement. These features can be summarized into three dimensions we hypothesize to underlie engagement in SimCityEDU including 1) Monitoring data 2) Focus on specific activities and 3) Sensitivity to Time.

We define the first dimension with measures of how long a player accesses the data maps concerning air pollution and power. Longer times likely reflect greater attention to monitor and maintain success in the game to reduce pollution and increase jobs. Since the data maps provide critical information for succeeding in the game, we hypothesize two motivations for accessing them including a) When a player has not met the challenge goal and therefore accesses the data maps to guide decisions about which actions to execute in order to increase jobs and reduce pollution. While not all engaged students may realize that it is necessary or helpful to study the data maps, it seems likely that studying the data maps indicates that a student is engaged. b) When a player has met the challenge goal but continues to refer to the data maps in order to maintain success. We include other variants of the data map feature such as total time and the total number of actions accessing the maps as compared to the average time and actions in the population of players.

For the second dimension, we include features that relate to a players frequency of actions for a given activity (e.g., bulldozing) to their total actions. Many of these features include measures that compare the number of actions for a given player to the average in the population. On this dimension, a disengaged player may hyper-focus on a given activity at levels that are much higher than what is observed in the population of players. A common example includes a student who nearly bulldozes their entire city. While fun to do, this type of excessive bulldozing often counters the goals of the challenge, and corresponds to the Without Thinking Fastidiously (WTF) construct proposed by Wixon and colleagues (2012).

Our third set of constructs is defined by features that aim to capture the amount of time a player focuses on a particular activity and comparing their time on an activity with the average in the population. Majority of these features capture how long a player engages in an activity with respect to total game time. For example, players with low levels of engagement may tend to spend little time with a tool and rapidly switch between different activities in a relatively short period of time.

We studied the relationship between learning and a set of 83 features. After completing the feature engineering, we followed a multi-step process to develop the model of cognitive engagement: feature selection and iterations through model fitting that best combined the features into a unified linear regression model using the M5-prime feature selection in RapidMiner 5.0 data mining software.

Model Results

Research Question 1: What are the key features of engagement related to game success among all player attempts?

Two algorithms were used to fit the detector of cognitive engagement, M5-prime and RepTREE (Witten & Frank, 2000). The best fitting model identified seven features of cognitive engagement correlated with game success including number of actions: zoning commercial, plopping alternative energy, and bulldozing coal and industrial. The other three features all relate to time spent on a given activity. These include total time bulldozing and time accessing the pollution and power maps. Both map features tap into our first dimension of engagement and helps to substantiate a piece of our original theory of engagement. Analyses showed that the M5-prime results in a more parsimonious model than RepTREE, with a student-level cross-validated correlation of .68. Correlation was computed using cross-validation (Efron & Gong, 1983), where the model is repeatedly trained on part of the data (in this case, some students) and tested on a held-out part of the data (in this case, other students).

Table 1. Model 1: Key Features of Player Engagement

N=153	Model 1: Attempt 1 | Any attempt
1	0.0027 * Total Zone Commercial
2	0.0199 * Plop (alternative)
3	0.0181 * Total BulldozeCoal
4	- 0.0032 * Total BulldozeInd
5	- 0.0004 * TTLTime_bulldoze
6	0.0016 * openedAirPollutionMap_TTLTime
7	0.003 * openedPowerMap_TTLTime
Correlation	**0.682**

Research Question 2: Among students who played at least twice, are the features of engagement the same after each game play?

The features of engagement extracted for Model 2 and Model 3 were found to predict game success. The prediction took the form of a linear regression and cross-validation was conducted to evaluate the detector's overall fit. As shown in the Table below, the number of features selected differ between Models therefore supporting our initial hypothesis. Model 2 results in 8 features and Model 3 reached 14 features. Model 3 results in 6 more features than Model 2 and there are five features common to both.

Table 2. Model 2 and Model 3 Key Features of Player Engagement

N=113	Model 2: Attempt 1 | At least 2 attempts	Model 3: Attempt 2 | At least 2 attempts
1	**0.0007 * Total Zone Commercial**	- 0.0002 * Actions
2	**0.0078 * Plop**	**+ 0.0411 * Plop**
3	**0.0025 * TTLTime_BulldozeCoal**	- 0.0017 * Dezone_comm
4	- 0.0012 * Map_Total Actions	- 0.0007 * Dezone_res
5	**0.0471 Map_TTLTime**	**0.0002 * Total Zone Commercial**
6	**0.0001 * TimeDuration**	0.0003 * Total Zone Industrial
7	0.0005 * TTLTime_BulldozeCom	**+ 0.0001 * TimeDuration**
8	+ 0.0006 * openedPopulation	- 0.0002 * Zone_total
9		+ 0.0066 * Coaldeselected
10		+ 0.003 * Coalviewhidden
11		- 0.0008 * Otherviewhidden
12		**+ 0.0042 * TTLTime_BulldozeCoal**
13		+ 0.0013 * closedAir Pollution
14		**+ 0.0414*Map_TTLTime**
Correlation	**0.698**	**0.715**

For both models, Models were trained separately on two groups of students using the cross-validation tool. The detector for Model 2 reached a cross-validated correlation coefficient of r=.698 while the detector for Model 3 achieved a correlation of r=.715.

Summary

Using data mining techniques, we were able to identify specific actions and patterns of play that suggest a player may be engaged. To do this, we fit three mathematical models that relates patterns of actions with success in meeting the game's objectives. Creating and modifying models of engaged behavior allowed us to test our hypotheses about the patterns of play that predict someone is engaged or disengaged with the given game mission "Jackson City". Each model identified those patterns of play that best predict when students are trying to meet the missions' goals as we designed them or some other set of goals that the player has chosen for herself. Now that we have initial models of engagement, we will iterate on the models by adding new features, using additional approaches to scoring or evaluating players' patterns of play and implement the early, rough version engagement detector itself in the game. As we work to improve the detector, we will also begin bringing the detector results to students and teachers to see how well it reflects their judgments about their own classes' play.

References

Brown, N. J., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review*, *23*(2), 221-234.

Efron, B.; Gong, G. (1983). A Leisurely Look at the Bootstrap, the Jacknife, and Cross-Validation. *The American Statistician*, 37 (1), 36 - 48.

Mislevy, R., Oranje, A., Bauer, M., Davier, V. A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K., & John, M. (2014). *Psychometric Considerations in Game-Based Assessment*. Retrieved from http://www.instituteofplay.org/work/projects/glasslab-research/.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, *55*(2), 503-524.

SimCityEDU [Computer software]. (2013). Institute of Play: GlassLab.

Witten and E. Frank. 2000. *Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan-Kaufmann, San Francisco, CA.

Wixon, M., Author, Gobert, J., Ocumpaugh, J. & Bachmann, M. (2012). WTF? Detecting students who are conducting inquiry without thinking fastidiously. *In Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization* (pp. 286-298).