

Design Space Sampling for the Optimization of Online Educational Games

Derek Lomas

HCI Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15212 USA
dereklovas@cs.cmu.edu

Erik Harpstead

HCI Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15212 USA
?@cs.cmu.edu

Abstract

Video game levels are configurations of a set of game parameters; the possible options represented by these game parameters constitute the total design space of a video game. By manipulating game parameters, level designers are able to craft and optimize player experiences. Our research involves the systematic exploration of the design space of an educational game in order to understand how different game parameters affect behavioral measures of learning and engagement. This position paper presents a variety of methods for procedurally generating level designs for the optimization of online educational games.

Author Keywords

Educational Games; Online Metrics; Learning; Engagement; Design Space

ACM Classification Keywords

J.m. Computer Applications: Miscellaneous

Introduction

Our research group is investigating a number line estimation game called "Battleship Numberline" (BSNL), a relatively popular online educational game. BSNL involves estimating the location of a number on a number line with two marked endpoints, such as estimating the location of the fraction $1/3$ on a number line from 0 to 1. Alternatively, players may be

presented with a marked position on a number line and have to estimate the numerical value of that position, such as in figure 1. The game arbitrarily represents this task in the context of a naval battle involving robotic pirates and forest animals. In classroom studies involving 20 minutes of gameplay, this game produced significant improvements in the estimation of fractions on a number line; moreover, it was positively viewed by over 90% of boys and 75% of girls [1].



Game Level Editor

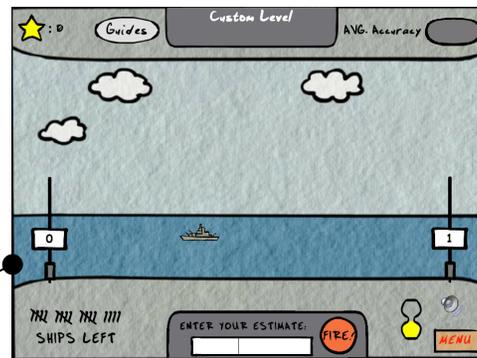


Figure 1: Battlehip Numberline game play

The BSNL level editor allows level designers to alter a variety of game parameters; specifically, there are 5 continuous variables (e.g. time limit or percent accuracy required for success) and 6 categorical parameters (e.g., choice of tickmarks or choice of decimals, fractions, whole numbers). These parameters can be configured for the purpose of creating specific units of instruction. For instance, a particular level of BSNL might be designed to teach the estimation of 2 digit decimal numbers on a number line from 0-1. Each parameter is expected to alter the nature of the challenge presented to the player, such that a

particular level design will require particular knowledge for successful performance. Furthermore, the nature of the challenge is expected to directly affect the player experience.

Online Educational Game Experiments

Because BSNL is currently played by several hundred players a day, we are able to conduct design experiments by randomly assigning players to different game conditions. These design experiments can potentially generate results that address theoretical issues in the psychology of learning or can be used for the practical optimization of the game's design for the purpose of maximizing our outcome metrics.

Outcome Metrics

Our research group is interested in creating effective educational games that are intrinsically motivating to play. Therefore, while BSNL collects log data about a wide range of player activity, our primary outcome metrics focus on player engagement and player learning. We operationally define player engagement in terms of intrinsic motivation, which can be measured behaviorally as the amount of time spent playing a game in a free choice setting [2]. Online game research commonly uses "time played" as a key metric of player satisfaction, though some researchers also consider the number of challenges completed (i.e., levels) as a corroborating measure [3]. Measuring learning within an online game poses substantial difficulties, particularly in situations when players cannot be tracked over time. We are currently using three different metrics to track learning over different game conditions: gain from an embedded pre-test to post-test, gain from early game performance to late game performance, and the learning curve documented by

performance over opportunities within a specific knowledge component.

Methods for Design Space Sampling

Theory Driven

The construction of game levels and level sequences may be based on an explicit theory of learning and gameplay. For instance, Bayesian Knowledge Tracing is a theoretically justified adaptive sequencing technique that is widely used in intelligent tutoring systems. We have designed game versions that utilize this design, which we can compare to a random sequence control. In other theory-driven design approaches, a single design variable can be adjusted. As an example, we can deploy versions of the game with and without time pressure to test the theory that time pressure supports the development of fluent number line estimation performance.

Example Driven

Game designs can also be produced to emulate existing game dynamics. For instance, we can construct game versions that mirror the common game design pattern of progressively increasing the difficulty of game levels.

Designer Driven

Individual game designers often have intuitive hunches about how to support learning and engagement. We've informally observed that different designers will produce surprisingly different game levels, even when pursuing the same goals. This suggests that a competition between game designers to produce effective levels (as measured by our game outcome metrics) may be a useful way of broadly sampling the game design space.

Fractional Factorial Designs

Response Surface Methodology (RSM) and fractional factorial designs are typically used in industrial manufacturing to optimize experimental designs involving large numbers of variables. However, we believe that these experimental design approaches may be useful for online game experiments, given the large number of experimental conditions that can be reasonably run. Both of these methods can be viewed as a "shotgun" approach that produces a large number of parameter configurations to discover optima. Fractional factorial designs are experimental techniques for maximizing the investigation of interaction effects between variables while minimizing the number of overall experimental conditions. While the results of such a design are inherently ambiguous, they can be used to screen for main effects and interactions of individual variables. RSM is a form of a fractional factorial experiment that can be used to predict optimal configurations of continuous variables. In a central composite design (a flavor of RSM), a designer can propose an optimal setting for each variable and provide an upper and lower bound of this variable. The experiment would then yield a set of datapoints that would allow for the prediction of optimal configurations, even if those were not explicitly tested.

Machine Learning Methods

The design space of games can also be explored through automatic machine learning methods. Work has been done employing genetic algorithms to evolve optimally engaging levels for simple platformer games that optimized for simple models of flow [4]. This work has been extended [5] to create a general framework that uses parameterized design elements within a genetic algorithm that can be constrained by arbitrary

optimization functions and applied to games of other genres.

Biographies

Derek Lomas is a design and learning science PhD student at the HCI Institute at Carnegie Mellon University. Lomas designs educational games that help support critical STEM skills in young learners. His game designs were recently awarded the "Impact Prize" by the United States Chief Technology Officer in the National STEM Game Competition. In 2009, Lomas received a MacArthur Foundation grant to support Playpower.org, an online community developing affordable, effective and fun learning games for underprivileged children around the world. Lomas received his MFA in Visual Arts from UC San Diego and his BA in Cognitive Science from Yale University, where he studied skill acquisition and the neuroscience of empathy.

References

- [1] Lomas D., Ching D., Stampfer, E., Sandoval, M., Koedinger, K. Battleship Numberline: A Digital Game for Improving Estimation Accuracy on Fraction Number Lines. *Conference of the American Education Research Association (AERA)* (2012)
- [2] Malone, T. Toward a theory of intrinsically motivating instruction. *Cognitive Science* 5, 4 (1981), 333-369.
- [3] Andersen, E., Liu, Y., Snider, R., Szeto, R., and Popovic, Z. Placing a value on aesthetics in online casual games. *CHI 2011, May 7-12, 2011, Vancouver, BC, Canada*, (2011).

Erik Harpstead is a first year PhD student at the HCI Institute at Carnegie Mellon University. Harpstead's work focuses on developing authoring tools for educational software including educational games and intelligent tutoring systems. He is currently a member of the DARPA ENGAGE project, where he works between Carnegie Mellon's teams at the Entertainment Technology Center and HCII to develop better methods and tools to create entertaining educational games for early elementary students. Harpstead received his BS in Psychology from Illinois Institute of Technology.

Acknowledgements

We thank all the developers and designers of Battleship Numberline.

- [4] Sorenson, N., & Pasquier, P. (2010). Towards a Generic Framework for Automated Video Game Level Creation. *Lecture Notes in Computer Science*, 6024, 131-140.
- [5] Sorenson, N., & Pasquier, P. (2010). The Evolution of Fun : Automatic Level Design through Challenge Modeling. *Proceedings of the First International Conference on Computational Creativity* (pp. 258-267).