Advances in Cognitive Systems X (20XX) 1–6 Submitted X/20XX; published X/20XX

Projective Replay Analysis: Using Cognitive Systems to Drive Evaluation of Educational Games

Erik Harpstead EHARPSTE@CS.CMU.EDU
Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15232 USA

Abstract

Understanding how an educational game behaves in the diverse contexts of individual player experience is an important element of educational game design. Capturing this understanding often requires expensive in-person playtests with specialized novice demographics in formal educational settings. To enable a quicker iterative design and evaluation process I propose Projective Replay Analysis, a technique for virtually playtesting a next iteration of an educational game using computational models to simulate the learning process of normal students. In my proposed work I plan to both demonstrate the technique of Projective Replay Analysis and validate whether virtual playesting paradigms, powered by computational models of player learning, reach similar conclusions to further expensive in-person tests.

1. Introduction

An important consideration to designing an educational game is to understand how the game reacts to different player experiences. Ideally, the game's instructional behaviors would align to the its instructional goals in order to foster players' learning of target content (Harpstead, MacLellan, Aleven, & Myers, 2014). While this may sound straightforward, educational games often involve complex dynamic elements (Hunicke, Leblanc, & Zubek, 2004) whose behaviors designers may not fully anticipate. This makes it essential to have a sense of how players go about exploring a game space and evaluate how the game reacts in the context created by players.

Gathering an understanding of players' behavior often requires some form of playtesting (Fullerton, 2014; Schell, 2008). In the typical playtesting session players from the target demographic are given the ability to play an early version of a game in order to gauge their reactions and assess the quality of the current iteration. When combined with analytics approaches (Loh, Sheng, & Ifenthaler, 2015; Seif El-Nasr, Drachen, & Canossa, 2013) these playtesting sessions can yield useful insights about the current state of a game's design and suggest useful future directions. In educational settings, however, playtesting approaches can become difficult because they require populations of content novices and often need to take place in formal education settings adding to administrative overhead.

In an effort to make the process of evaluating and iterating on educational games easier I propose to demonstrate the Projective Replay Analysis approach. This approach uses replay fidelity log traces of playtesting sessions as training data to computational models of human learners that can then be used to do first-pass evaluations of new game design iterations without having to gather further human data. In developing this approach, it is necessary to create a computational model that is both capable of dealing with the complexities of a game environment while also modeling the knowledge acquisition process of novices.

2. Projective Replay Analysis

Replay Analysis is an approach I developed to enable several different game analytics techniques using a single base of recorded data from an educational game (Harpstead, MacLellan, Aleven, & Myers, 2015). Central to this approach was the use of replay fidelity log files (as opposed to simple recorded metrics) that can be played back through the game engine. The live game state in the engine can then be used to yield various measures of the player experience. These replay files are similar to the ones commonly seen in strategy games that have been used in work on strategy detection (Weber & Mateas, 2009).

The original replay analysis approach consisted of two main components: a particular schema for logging player actions, and a Replay Analysis Engine (RAE) for taking recorded actions and re-enacting them through the game engine (Harpstead, Myers, & Aleven, 2013). The logging schema is used to capture player actions as well as relevant game context at the level of a basic action, defined as the smallest unit of meaningful action that a player can exert on the game. These actions are meant to be contextualized to the game world (e.g., picking up or dropping an object), rather than the raw input of the player (e.g., mouse down at position (x, y)). Additionally, each action is paired with a description of the state of the game just before the action took place to provide contextual information. The paired recording of state is important in situations where a game's state and behavior could change for reasons other than direct player action (e.g., a physics engine simulating the motion of objects, or a non-player character making its own independent decisions).

The second major component of the approach is the Replay Analysis Engine (RAE). The RAE reads in a player's log file and reconstructs the player's play session action-by-action. For each action, the RAE first constructs the state in which the action took place and then enacts the player action to let the game engine resolve any consequences of the action, using the same code that would normally handle such an action. Analyses can then be performed by augmenting the replayed state to create new measures with full access to any state attributes that would have been present at playtime. These analyses represent an accurate reproduction of the player's own experience because the re-instantiated state is composed of exactly the same game elements, in terms of code. Having paired states with each action also allows logs to be replayed accurately without having to interpolate prior actions.

Projective Replay Analysis extends the replay concept by enabling historically recorded replay files to be applied to new versions of the game. This requires the addition of a third component, an agent module for handling the ambiguities that arise between the old and new versions of a game. In my proposed work this component will take the form of a computational player model used to simulate the decision-making processes of playtesters. I plan to explore two forms of this player model Literal Replay, and Flexible Replay. In the literal form, the player model simply acts as the normal RAE in the new game context by re-enacted players' actions as they occurred in the replay file. If literal replay ever encounters states or actions that are no longer compatible with new game mechanics it will simply fail and move on.

The second form of player model is one augmented with an AI design making process. This player model takes demonstrations from players' replays and learns to perform its own actions within the game, in a style similar to the respective player. The core of the flexible player model is the TRESTLE algorithm, a model of human concept formation (Maclellan, Harpstead, Aleven, & Koedinger, 2015) that learns hierarchical concept trees given structured examples. Using an Apprentice Learner Architecture paradigm (Maclellan, Harpstead, Patel, & Koedinger, 2016), an action planner can be used to translate demonstrations into generalized action sequences and TRESTLE can be used to learn concepts corresponding to when those action sequences should be employed. These learned action concepts can then be used to perform model tracing similar to how intelligent tutoring systems employ production rules (Aleven, 2010).

3. Future Work

In my prior work I have applied Replay Analysis to an existing version of the educational game *RumbleBlocks* (Christel et al., 2012; Harpstead et al., 2015). This work has yielded several design recommendations to the game that have the potential to improve the alignment between its instructional behavior and its educational goals (Harpstead et al., 2014). To validate whether these recommendations were correct I plan to employ Projective Replay Analysis. Additionally, I will seek to validated whether virtual playtesting paradigms reach the same conclusions as new in-person testing. To do this I plan to undertake three studies.

The first study will apply a Literal replay paradigm to the suggesting game design variations. This study will first demonstrate the limitations of the using old data in a novel game and establish a baseline for how drastically the different mechanical variations have affected the possible solutions to in-game challenges.

The second study will employ flexible replay to the same variations explored in study one. To explore this question, instead of replaying student's literal logs in the newly designed game variations, I will use them as training data for a player model. I will create an individual player model for each of the players within my dataset. This player model will receive the student's actions as demonstrations, which it will use to learn skill concepts for the game. The player models will then be used to play the game variations in order to generate new game actions. As the agents interact with the new game system they will incorporate any feedback from their respective successes and failures to further refine their models. The solutions created by these actions will then be evaluated as if they had been generated by the students themselves. Ideally these solutions will be similar in character to the ones the students originally created, but they will have a chance to be distinct and responsive to the new dynamics created by the variation in game mechanics.

The final study will be a full classroom evaluation of the game variations using real students. The data gathered in this final study will allow me to both compare the overall conclusions of the game variations with regard to design quality, and evaluate how the particular behaviors of the virtual agents compare to a new group of real people. This aspect of the work will stand to contribute to the cognitive systems community by directly evaluating the performance of novice learner models and real human novices. I suspect the results of this work will not only enable the development or more efficient playtesting paradigms for educational games but also provide a useful testbed for advancing our understand of cognitive systems more broadly.

Acknowledgements

This work was supported in part by the DARPA ENGAGE research program under ONR Contract Number N00014-12-C-0284 and by a Graduate Training Grant awarded to Carnegie

Mellon University by the Department of Education # R305B090023. All opinions expressed in this article are those of the authors and do not necessarily reflect the position of the sponsoring agency.

References

Aleven, V. (2010). Rule-Based Cognitive Modeling for Intelligent Tutoring Systems. In *Advances in Intelligent tutoring Systems* (pp. 33–62).

Christel, M. G., Stevens, S. M., Maher, B. S., Brice, S., Champer, M., Jayapalan, L., … Lomas, D. (2012). RumbleBlocks: Teaching science concepts to young children through a unity game. In *Proc. CGAMES 2012* (pp. 162–166). IEEE. http://doi.org/10.1109/CGames.2012.6314570

Fullerton, T. (2014). *Game Design Workshop*. Boca Raton, FL: CRC Press.

Harpstead, E., MacLellan, C. J., Aleven, V., & Myers, B. A. (2014). Using extracted features to inform alignment-driven design ideas in an educational game. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 3329–3338). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557393

Harpstead, E., MacLellan, C. J., Aleven, V., & Myers, B. A. (2015). Replay Analysis in Open-Ended Educational Games. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 381–399). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-05834-4_17

Harpstead, E., Myers, B. A., & Aleven, V. (2013). In search of learning: facilitating data analysis in educational games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 79). New York, New York, USA: ACM Press. http://doi.org/10.1145/2470654.2470667

Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA : A Formal Approach to Game Design and Game Research. In *Proc. of the AAAI Workshop on Challenges in Game AI* (pp. 1–5).

Loh, C. S., Sheng, Y., & Ifenthaler, D. (Eds.). (2015). *Serious Games Analytics*. Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-05834-4

Maclellan, C. J., Harpstead, E., Aleven, V., & Koedinger, K. R. (2015). TRESTLE: Incremental Learning in Structured Domains using Partial Matching and Categorization. In *Proceedings of the 3rd Annual Conference on Advances in Cognitive Systems - ACS 2015* (pp. 28–31). Atlanta, GA: Cognitive Systems Foundation.

Maclellan, C. J., Harpstead, E., Patel, R., & Koedinger, K. R. (2016). The Apprentice Learner Architecture: Closing the loop between learning theory and educational data. In *Proceedings of the 9th International Conference on Educational Data Mining - EDM '16* (p. to appear).

Schell, J. (2008). *The Art of Game Design: A Book of Lenses* (1st ed.). Burlington, MA: Morgan Kaufmann.

Seif El-Nasr, M., Drachen, A., & Canossa, A. (Eds.). (2013). *Game Analytics*. London: Springer London. http://doi.org/10.1007/978-1-4471-4769-5

Weber, B. G., & Mateas, M. (2009). A data mining approach to strategy prediction. In *2009 IEEE Symposium on Computational Intelligence and Games* (pp. 140–147). Ieee. http://doi.org/10.1109/CIG.2009.5286483