Standardizing the Evaluation of Digital
Managers for Better Interactive Experiences

Giulio Mori

Department of Computer Science
Reykjavik University
101 Reykjavik, Iceland
giulio17@ru.is

Abstract

As Artificial Intelligence (AI) technologies become increasingly integrated with society (e.g., in the form of digital assistants, games, or self-driving cars), it becomes increasingly important to understand how they can ensure that people's experiences meet certain desirable criteria. In Experience Management, researchers and professionals seek to create automated systems called "experience managers", which work to improve people's experiences in designer-specified ways. Unfortunately, the evaluation and comparison of experience managers is currently impaired by the lack of both a common platform to evaluate them and a common language to describe experience management tasks. Experience in other fields such as General Game Playing and AI Planning has shown that having a common platform can stimulate and accelerate research progress. The goal of this project is thus to develop a platform for evaluating and comparing arbitrary experience managers on a wide variety of tasks, to accelerate the essential progress of Experience Management research. We plan to do this by studying existing experience managers, developing a common platform that can support them, and promoting the platform to the research community as a tool for their research. Throughout our work, we will draw lessons from the field of General Game Playing, hoping to reuse some of its existing infrastructure and stimulate new exchanges of ideas between the two fields of research.

1 Introduction

Experience Management is a subfield of Artificial Intelligence (AI) that investigates automated systems called "experience managers", which serve to enhance people's experiences in a digital application according to one or more given metrics. The research of Experience Management has its origins in Drama Management (Laurel 1986; Weyhrauch 1997) and Intelligent Tutoring Systems (Sleeman and Brown 1982; Koedinger et al. 1997). Most lines of recent academic work in Experience Management began as a generalization of Drama Management, which attempts to ensure that dramatically relevant stories occur in an interactive, narratively rich environment (e.g., a story-based video game). The term "Experience Management" was first

coined by Riedl et al. (2008). Although a significant amount of experience managers have been built over the past two decades, the study of Experience Management is currently impaired by two fundamental lacks: it has no standard language for describing experience management tasks, and it has no comprehensive platform for evaluation and comparison. The outcome is that research in the field is widely fragmented and mostly independent, making it tough to draw conclusions from research that generalize well across the field (Koenitz et al. 2011; Szilas, Boggini, and Petta 2011; Szilas et al. 2012). Some prior work has offered structural comparisons of experience managers (e.g., in terms of their design and expected capacities) (Roberts and Isbell 2008; Thue 2015) and different work has compared one manager's performance to that of a homemade baseline (Thue et al. 2007; Ramirez and Bulitko 2014; Hernandez, Bulitko, and Hilaire 2014). We are unaware of any previous efforts to directly compare the performance of two or more independently-created managers. To address these needs, I aim to build both a general language for expressing experience management tasks and a general platform for evaluating and comparing experience managers.

To help with this task, I will draw inspiration from the subfield of AI "General Game Playing " (GGP) (Genesereth and Love 2005) whose goal is to find methods to play games well that are independent from the game being played. In General Game Playing, the Game Description Language (GDL) (Love, Genesereth, and Hinrichs 2006) was defined to examine agents' abilities to play games and make them compete against each other. GDL is a logic-based language that allows representing the rules of a game (legal moves, state transitions, goals of the players, etc.) in a declarative way. GDL allows the players to interpret the rules of a game both to simulate the game as well as to analyze the game for finding structures that might help develop a successful strategy for playing the game. In its original form, GDL is limited to discrete, deterministic, perfect information games with an arbitrary number of players. However, there exist extensions for non-determinism, imperfect information (Schiffel and Thielscher 2014), and even introspection of the players' knowledge (Thielscher 2017).

Games are frequently employed as valuable testing

grounds for AI technologies, as their interactive characteristics make them useful as simulations of the real world. The shared challenge of making good decisions in a given environment (such as a game, a training simulation, or the real world) is the first of various similarities that the fields of GGP and Experience Management share. Both fields work with environments that vary over time according to specific rules, and those variations are determined by the actions of multiple agents that have potentially different and conflicting goals. The goal of General Game Playing is to find game-independent methods to play games well, while Experience Management aims to find general methods to drive experiences well.

2 Proposed Research

Over the past decade, the study of GGP has benefited from the development of a common language (Thielscher 2011a) and a common evaluation platform, both in terms of an increase in publications and a better shared understanding in the field (Thielscher 2011b). Given that, as stated before, the GGP field and the Experience Management field share many similarities: can common ground be established between the challenges of GGP and Experience Management such that each field could benefit from solutions found in the other? I plan to study this question, and also provide a language and platform that support evaluating and comparing experience managers. Another question is, what guarantees can be established regarding the behaviour of an experience management system, and what assumptions would they require? Answering this question will require a formal definition of the experience management task and the capabilities of the system.

Overall, my research goals can be summarized as follows:

- to develop a formal language for describing environments for experience management tasks;
- to develop a platform that connects experience managers with an environment and supports evaluating experience managers according to different metrics;
- to promote the use of the developed platform to stimulate progress in research of experience management.

By learning from the experiences of researchers in the field of General Game Playing (both good and bad) in their highly similar pursuit of standardization, I will improve my chances of avoiding similar pitfalls and achieving similar success.

Figure 1 outlines the approach that I plan to follow for creating a platform that can support such comparative evaluations. To be able to compare how two different managers perform in an arbitrary given environment, I have started to investigate the requirements of the evaluation platform (WP 1). Next, I will develop a language for specifying environments that respect the platform's established requirements (WP 2). In WP 3, I will develop the platform itself, which I will evaluate using an example environment and at least two experience managers in WP 4. Descriptions of each work package and task are outlined below.

WP 1 - Evaluation Platform Requirements. The goal of this work package is to establish the requirements for a
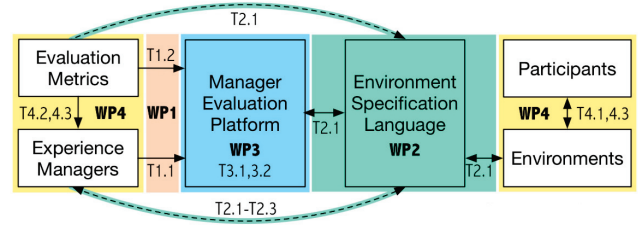

Figure 1: A diagram of the proposed architecture of my evaluation platform and its related elements. Each shaded area denotes one of my first four work packages (WP#). Many tasks correspond to one or more arrows in the diagram (T#.#). WPs 1 through 3 cover the core of my approach, while WP4 covers my evaluation of its success.

platform that can support the evaluation of a variety of different experience managers. This goal can be achieved by three tasks: (T1.1) study and summarize methods for Experience Management AI to allow an arbitrary manager to be evaluated using our platform, (T1.2) study and summarize potential evaluation methods to support the comparison of different managers, and (T1.3) establish requirements for an evaluation platform. At the time of writing this paper, I have finished the first task, and I am currently investigating evaluation methods for Experience Management.

WP 2 - Environment Specification Language. The goal of this work package is to develop a language to specify general experience management tasks and understand how different uses of the language can influence a manager's behaviour. Like in WP 1, this work package is divided in three tasks: (T2.1) develop a language to specify general experience management tasks, (T2.2) learn which interactions can avoid undesirable behaviour, and (T2.3) learn which interactions are necessary for desired behaviour.

WP 3 - Implementing the Platform. The goal of this work package is to develop, document and promote the manager evaluation platform. This work package is also divided into three different tasks: (T3.1) design the architecture of the platform, (T3.2) implement the platform, and (T3.3) document and promote the platform.

WP 4 - Evaluation of Language and Platform. The goal of this work package is to assess the value of our environment specification language and manager evaluation platform by comparing managers in an environment specified using our language. The work that needs to be done to achieve this objective is also divided in three tasks: (T4.1) choose and specify an environment to prove the concept of our approach, (T4.2) identify and implement/modify at least two experience managers to compare different managers inside the environment developed, and (T4.3) compare the implemented managers in the specified environment to assess the value of the approach developed.

References

Genesereth, M., and Love, N. 2005. General game playing: Overview of the aaai competition. *AI Magazine* 26:62–72.

Hernandez, S. P.; Bulitko, V.; and Hilaire, E. S. 2014. Emotion-based interactive storytelling with artificial intelligence. In *Proceedings of the Tenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE'14, 146–152. AAAI Press.

Koedinger, K. R.; Anderson, J. R.; Hadley, W. H.; and Mark, M. A. 1997. Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)* 8:30–43.

Koenitz, H.; Haahr, M.; Ferri, G.; and Sezen, T. I. 2011. Towards a unified theory for interactive digital storytelling - classifying artifacts: A workshop at icids 2011. In *Proceedings of the 4th International Conference on Interactive Digital Storytelling*, ICIDS'11, 360–361. Berlin, Heidelberg: Springer-Verlag.

Laurel, B. K. 1986. *Toward the Design of a Computer-Based Interactive Fantasy System*. Ph.D. Dissertation, Ohio State University.

Love, N.; Genesereth, M.; and Hinrichs, T. 2006. General game playing: Game description language specification. Technical Report LG-2006-01, Stanford University, Stanford, CA. http://logic.stanford.edu/reports/LG-2006-01.pdf.

Ramirez, A., and Bulitko, V. 2014. Automated planning and player modeling for interactive storytelling. *IEEE Transactions on Computational Intelligence and AI in Games* 7(4):375–386.

Riedl, M. O.; Stern, A.; Dini, D. M.; and Alderman, J. M. 2008. Dynamic Experience Management in Virtual Worlds for Entertainment, Education, and Training. *International Transactions on Systems Science and Applications - ITSSA* 4.

Roberts, D. L., and Isbell, C. L. 2008. A survey and qualitative analysis of recent advances in drama management. *International Transactions on Systems Science and Applications - ITSSA* 4(2):61–75.

Schiffel, S., and Thielscher, M. 2014. Representing and reasoning about the rules of general games with imperfect information. *J. Artif. Int. Res.* 49(1):171–206.

Sleeman, D., and Brown, J. S. 1982. *Intelligent Tutoring Systems*. London : Academic Press.

Szilas, N.; Rank, S.; Petta, P.; and Müller, W. 2012. Sharing interactive digital storytelling technologies ii. In Oyarzun, D.; Peinado, F.; Young, R. M.; Elizalde, A.; and Méndez, G., eds., *ICIDS*, volume 7648 of *LNCS*, 216. Heidelberg / Dordrecht / London / New York: Springer.

Szilas, N.; Boggini, T.; and Petta, P. 2011. Sharing interactive digital storytelling technologies. In *Proceedings of the 4th International Conference on Interactive Digital Storytelling*, ICIDS'11, 366–367. Berlin, Heidelberg: Springer-Verlag.

Thielscher, M. 2011a. The general game playing description language is universal. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, 1107–1112. AAAI Press.

Thielscher, M. 2011b. General game playing in ai research and education. In Bach, J., and Edelkamp, S., eds., *KI 2011: Advances in Artificial Intelligence*, 26–37. Berlin, Heidelberg: Springer Berlin Heidelberg.

Thielscher, M. 2017. Gdl-iii: A description language for epistemic general game playing. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 1276–1282. AAAI Press.

Thue, D.; Bulitko, V.; Spetch, M.; and Wasylishen, E. 2007. Interactive storytelling: A player modelling approach. In *3rd AAAI Conference on AI and Interactive Digital Entertainment*, AIIDE'07, 43–48. AAAI Press.

Thue, D. J. 2015. *Generalized Experience Management*. Ph.d. dissertation, University of Alberta.

Weyhrauch, P. 1997. *Guiding interactive drama*. Ph.d. dissertation, Canergie Mellon University, Pittsburgh, PA.