Towards Lifelong Interactive Learning For Open-Ended Co-creative Embodied Narrative Improvisation

Mikhail Jacob
Georgia Institute of Technology
Atlanta, GA 30308 USA
mikhail.jacob@gatech.edu

Abstract

This article describes a doctoral research plan for modeling the co-creative improvisation of open-ended embodied narratives in two ongoing research projects called LuminAI and GENII. The research focuses on mitigating the knowledge-authoring bottleneck involved using lifelong interactive learning from observation and experience. The research methodology involves building co-creative systems for human - computer embodied narrative improv, public installation/exhibition/dissemination in order to collect real-world datasets of human-computer improv, human-centered evaluation of the experience in public/lab settings, and evaluation of the agent learning/performance using the collected data sets. The article concludes with a research timeline describing the ongoing work.

Introduction

The computational creativity community has formally studied creative cognition in various domains (c.f. Gardner 2008), while also attempting to model creativity computationally (Veale, Feyaerts, and Forceville 2013). More recently there has been an increased focus on co-creativity between humans and computers. However, the popularity of virtual reality (VR) as an interactive storytelling medium has emphasized the lack of computational models of co-creativity in embodied creative domains such as embodied narrative improvisation (embodied narrative improv).

VR games and experiences are a fast-growing segment of the digital entertainment (BUSINESS WIRE 2017) as well as training and simulation industries (Cooke and Stone 2013; Seymour et al. 2002). Naturalistic interaction in VR, such as walking and manipulating the virtual world using the body, supports the open-endedness of improv experiences. Users can physically mime flipping burgers (Owlchemy Labs 2016) or scaling Mt. Everest (Sólfar Studios 2016) to do so in-game. In contrast, non-player characters (NPCs) are limited to using small sets of canned animations due to the knowledge-authoring bottleneck (Csinger, Booth, and Poole 1994) involved in expanding this set. This results in a jarring imbalance in the creative autonomy available to (and responsibility placed on) the human collaborator in such an embodied co-creative experience. Thus the knowledge-authoring bottleneck is a significant challenge to developing computational models of co-creativity in open-ended embodied narrative improv.

Various techniques exist in the game AI literature to mitigate the knowledge-authoring bottleneck. However, the flexibility and expressivity of interaction required for embodied narrative improv makes them less applicable. NPC movement can be learned/generated from motion capture datasets. However, few datasets exist for a broad enough set of human action to enable open-ended improvisation. Additionally, few (if any) motion capture data sets are annotated with formal semantics describing how the actions within them affect their environment in order to choose when to use them. Procedurally generated movements can also appear unnatural, exhibiting the uncanny valley effect (Mori, MacDorman, and Kageki 2012).

One way to mitigate the knowledge-authoring bottleneck in an open-ended improv system is to learn the necessary knowledge from all the human users over the system's lifetime. This would be facilitated by a knowledge representation that could be updated incrementally, integrating new and prior learned knowledge in an online manner. Additionally, in order to make the learned knowledge usable immediately, the user could interactively correct learning errors.

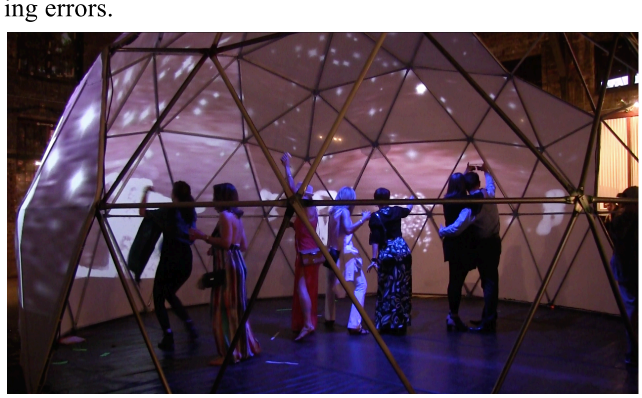

Figure 1: LuminAI dome installation with people improvising open-ended embodied proto-narratives with each other and virtual characters (humanoid figures made of motes of light)

This article describes a doctoral research plan for modeling the co-creative improvisation of open-ended embodied narratives in two ongoing research projects called LuminAI and GENII. The research focuses on mitigating the knowledge-authoring bottleneck involved using lifelong interactive learning from observation and experience. The research methodology involves building co-creative systems for human - computer embodied narrative improv, public installation/exhibition/dissemination in order to collect real-world datasets of human-computer improv, human-centered evaluation of the experience in public/lab settings, and evaluation of the agent learning/performance using the collected data sets. The article concludes with a research timeline describing the ongoing work.

LuminAI

LuminAI (formerly *Viewpoints AI*) is a dome-based interactive art installation (Jacob et al. 2013) where human participants can interact expressively with virtual characters in order to co-creatively improvise movement-based performances together. Over its lifetime, LuminAI learns in real-time what actions it can perform and how to sequence them in response to a collaborator's actions (Jacob and Magerko 2015) by observing their movements. Learned actions can be flexibly reused in many different contexts. The system also learns how to sequence actions together by learning Markov chains (Cope and Mayer 1996) of action classes from observation. This models the simplified improv process as a mutual coupling where the characters and users are responding to each other.

In addition, the installation utilizes a computational formalization of Viewpoints movement theory (Bogart and Landau 2006) to analyze the improvised scene, organize its experiences, and aesthetically transform learned actions for variety (Jacob and Magerko 2015). The system models improvisational responses by choosing responses using multiple response strategies within a cognitive architecture called Soar (Langley, Laird, and Rogers 2009) that performs the action selection reasoning. These strategies include repetition (to create rapport), transformation along Viewpoints dimensions (to create novelty), retrieval of previously observed movements from episodic memory (Tulving 1985) that were similar in Viewpoints dimensions (to generate aesthetically appropriate novelty), and using the learned Markov chains of action classes discussed before (to follow expected patterns/develop motifs).

LuminAI has been exhibited publically in informal, invited, and peer-reviewed showings. An iterative design methodology was used, with participant feedback incorporated after each showing. The installation has also been formally evaluated using theoretical analysis of authorial leverage and ablative user study (Jacob, Zook, and Magerko 2013) as well as qualitative study during a public exhibition (Long et al. 2017).

Using a working definition of narrative as a sequence of causally and temporally related actions performed by a set of characters, the improvised performance in LuminAI constitutes a proto-narrative. This is an initial step along the stated research direction because the actions performed in it are temporally and aesthetically related but not fully causally related. Additionally, the improvisation focused on pure abstract movement, ignoring the environmental context of those movements. The ongoing MImE/GENII project addresses these issues.

MImE/GENII

MImE (Movement-based Improvisation Environment) is a VR experience where a person and a virtual character can collaboratively improvise open-ended embodied narratives. The initial research prototype being built enables embodied narrative improv within a larger framing narrative of the protagonist and their service android crash landing on an alien planet, damaging the android's memory. In order to survive, explore, and escape, the player and their android have to perform collaborative tasks using various objects in the environment with the player teaching the android the necessary skills to help complete the task. A future version with less structure is planned for free improvisation of embodied narrative, but is out of scope for the current research plan.

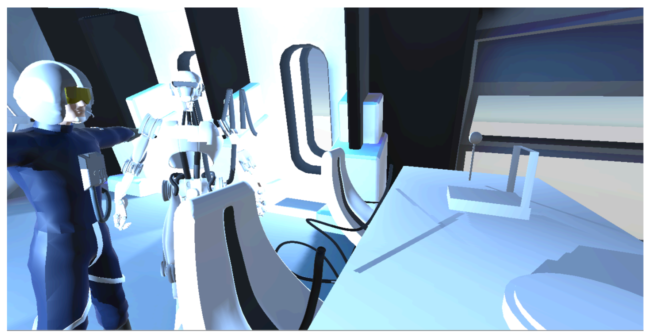

Figure 2: The MImE virtual environment (screenshot) for co-creating improvised embodied narratives about marooned space explorers and their robots.

GENII (Graphical Embodied Narrative Intelligence & Improvisation) is the intelligent agent architecture for lifelong learning and virtual character control in open-ended embodied narrative experiences such as MImE. In order for the GENII-controlled android character in MImE to learn how to interact with the environment, it observes the player's interactions with it, reasons about how the player's physical body movements change the state of objects in the environment and uses that to build an inductive model of the effects of the player's action. It uses this model to perform an action in the environment during collaborative tasks with feedback from the player. GENII agents thus perform interactive imitation learning (Mühlig, Gienger, and Steil 2012) of action models (Yang, Xu, and Chen 1997).

The learned action models are stored in a generalized hypergraph (Iordanov 2010) knowledge representation. A generalized hypergraph is an extension of hypergraphs and graphs where single edges can connect an arbitrary number of vertices and edges themselves. This allows GENII to hierarchically connect the grounded percepts it experiences to abstract concepts that are derived from them.

The hypergraph implements a cognitive memory model that combines sensory, episodic, and semantic memories into a single structure. The percepts the agent receives are temporarily stored in sensory memory, while a reconstructive representation of the percept is then stored in the hypergraph linked to the last and next percepts the agent experiences. This forms the agent's episodic memory. These percepts are grounded in the agent's experience and are executable action models. The agent clusters over similar episodes or actions in order to learn a prototypical action concept or class. These action concepts can then be referenced independent of the individual episodes in which they were learned. This implements the agent's semantic memory of actions.

The knowledge that is stored at each layer of the hypergraph grows more abstract in the vertical direction. With repeated examples of actions, the knowledge can be clustered horizontally at each layer of the knowledge hierarchy. Particularly, sequences of human position data observed by the agent are stored at the lowest layer of the hypergraph. This positional data is normalized and generalized for flexible usage.

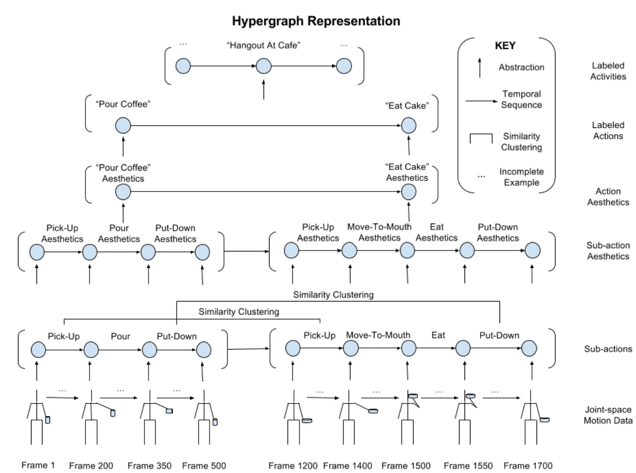

Figure 4: A partial example of the hypergraph knowledge representation

Each sequence of motion data is segmented at the next layer by the changes they cause to the environment. This world state is described using image schematic primitives such as containment (whether an object contains any other object or is itself contained by another object), linkage (whether an object is physically linked to other objects), center-periphery spatial relationships (whether some objects encircle others at some distance), joining, splitting, etc. Changes in the state of environmental objects are tracked. An embodied approach is applied to limit the tracking to objects the user interacts with (this limits the second order effects learnable but restricts change tracking to a manageable problem). The movements between each change in the environment are referred to as sub-actions.

With repeated examples, sub-actions are clustered together based on similarity of motion and effect (on world state). Labels for prototypes of each sub-action cluster are then automatically crowdsourced using Mechanical Turk. This is required to enable communication of learned knowledge to human users as well as to enable future connection to symbolic knowledge bases. These labels are then cleaned and processed. This crowdsourcing is repeated for the action as well, which is composed of a sequence of sub-actions. Given a goal state, this hierarchical action model learned by the agent can be used to act in the world through action planning (Erol, Hendler, and Nau 1994). It can also be connected to knowledge bases for common sense (Speer and Havasi 2012) and linguistic reasoning (Martha, Dan, and Paul 2005), or used for analogical reasoning (Gentner and Smith 2012) and conceptual blending (Fauconnier and Turner 2008) in the future, but that is out of scope for this research.

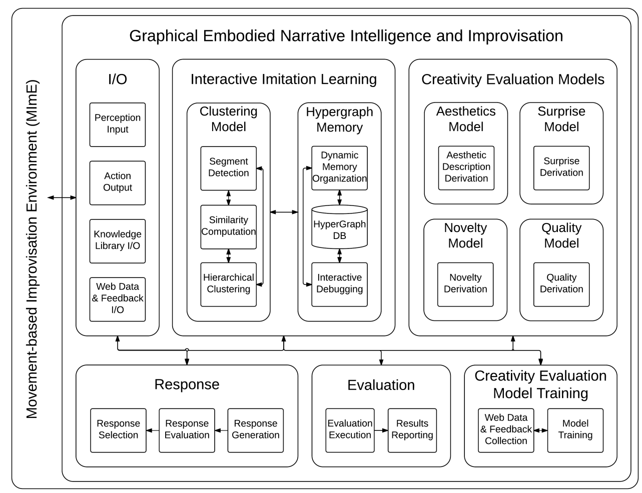

Figure 3: The GENII agent architecture.

Another key component of GENII is a set of metacognitive (Flavell 1979) models of creativity that evaluate the novelty, value, surprise, and aesthetic description of knowledge (like in Maher 2010) at each layer of the hypergraph. This is used to filter out low quality knowledge from the agent. This includes ignoring bad percepts during interactive learning, forgetting action models during system idle time, and generating only appropriately creative responses from sequences or combinations of learned knowledge.

Research Plan

After initial investigations into embodied narrative improv with the LuminAI project, the author's focus is now on the MImE/GENII project that is ongoing with an initial prototypes being completed for the start of summer 2017. At the time of writing, there is one partial user task completed in MImE with associated assets. In GENII, the motion layer learning has been added and evaluated along with a model of motion dissimilarity. By summer, MImE will include 2-3 open-ended user tasks within the crashed spaceship interior. In GENII, this will include initial models of novelty and value; tracking of partial object state; and motion layer segmentation using object state changes. This will allow baseline user evaluations and system evaluations in summer 2017. With results and feedback from them, in fall 2017, the remaining MImE scenario will be added (base camp). In GENII, novelty, surprise, value, and aesthetic description models; full object state tracking; and response generation with learned action models will be completed. Spring 2018 will be spent evaluating the system, refining it, and writing the dissertation. The defense will be in summer 2018.

Acknowledgements

I would like to thank my advisor, Dr. Brian Magerko, for his support and the rest of the ADAM Lab at Georgia Tech for their help.

References

Bogart, A., and T. Landau. 2006. *The Viewpoints Book: A Practical Guide to Viewpoints and Composition*. Theatre Communications Group.

Cooke, Neil, and Robert Stone. 2013. "RORSIM: A Warship Collision Avoidance 3D Simulation Designed to Complement Existing Junior Warfare Officer Training." *Virtual Reality* 17 (3): 169–79.

Cope, David, and Melanie J Mayer. 1996. *Experiments in Musical Intelligence*. Vol. 12. AR editions Madison, WI.

Csinger, Andrew, Kellogg S Booth, and David Poole. 1994. "AI Meets Authoring: User Models for Intelligent Multimedia." *Artificial Intelligence Rev.* 8 (5-6): 447–68.

Erol, Kutluhan, James Hendler, and Dana S Nau. 1994. "HTN Planning: Complexity and Expressivity." In *AAAI*, 94:1123–28.

Fauconnier, Gilles, and Mark Turner. 2008. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

Flavell, John H. 1979. "Metacognition and Cognitive Monitoring: A New Area of Cognitive–developmental Inquiry." *American Psychologist* 34 (10): 906.

Gardner, Howard. 2008. *Art, Mind, and Brain: A Cognitive Approach to Creativity*. Basic Books.

Gentner, Dedre, and Linsey Smith. 2012. "Analogical Reasoning." *Encyclopedia of Human Behavior* 130: 130.

Iordanov, Borislav. 2010. "HyperGraphDB: A Generalized Graph Database." In *International Conference on Web-Age Information Management*, 25–36. Springer.

Jacob, Mikhail, Gaëtan Coisne, Akshay Gupta, Ivan Sysoev, Gaurav Verma, and Brian Magerko. 2013. "Viewpoints AI." In *Proceedings of the 9th Artificial Intelligence and Interactive Digital Entertainment Conference*.

Jacob, Mikhail, and Brian Magerko. 2015. "Interaction-Based Authoring for Scalable Co-Creative Agents." In *Proceedings of International Conference on Computational Creativity*.

Jacob, Mikhail, Alexander Zook, and Brian Magerko. 2013. "Viewpoints AI: Procedurally Representing and Reasoning about Gestures." In *Proceedings of the Digital Games Research Association*.

Labs, Owlchemy. 2016. *Job Simulator: The 2050 Archives*. Owlchemy Labs.

Langley, Pat, John E Laird, and Seth Rogers. 2009. "Cognitive Architectures: Research Issues and Challenges." *Cognitive Systems Research* 10 (2): 141–60.

Long, Duri, Mikhail Jacob, Nicholas M. Davis, and Brian Magerko. 2017. "Designing for Socially Interactive Systems." To appear in *Proceedings of the Eleventh ACM Conference on Creativity and Cognition*, 10.

Maher, Mary Lou. 2010. "Evaluating Creativity in Humans, Computers, and Collectively Intelligent Systems." In *Proceedings of 1st DESIRE Network Conference on Creativity and Innovation in Design*, 22–28. Desire Network.

Martha, Palmer, Gildea Dan, and Kingsbury Paul. 2005. "The Proposition Bank: A Corpus Annotated with Semantic Roles." *Computational Linguistics Journal* 31 (1).

Mori, Masahiro, Kari F MacDorman, and Norri Kageki. 2012. *The Uncanny Valley: The Original Essay by Masahiro Mori*.

Mühlig, Manuel, Michael Gienger, and Jochen J Steil. 2012. "Interactive Imitation Learning of Object Movement Skills." *Autonomous Robots* 32 (2): 97–114.

Seymour, Neal E, Anthony G Gallagher, Sanziana A Roman, Michael K O'Brien, Vipin K Bansal, Dana K Andersen, and Richard M Satava. 2002. "Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study." *Annals of Surgery* 236 (4): 458–64.

Speer, Robert, and Catherine Havasi. 2012. "Representing General Relational Knowledge in ConceptNet 5." In *LREC*, 3679–86.

Studios, Sólfar. 2016. *Everest VR*. Sólfar Studios.

Tulving, Endel. 1985. "Elements of Episodic Memory."

Veale, Tony, Kurt Feyaerts, and Charles Forceville. 2013. *Creativity and the Agile Mind: A Multi-Disciplinary Study of Multi-Faceted Phenomenon*. Vol. 21. Walter de Gruyter.

WIRE, BUSINESS. 2017. *Worldwide Spending on Augmented and Virtual Reality Forecast to Reach $13.9 Billion in 2017, According to IDC*. http://www.businesswire.com/news/home/20170227005097/en/Worldwide-Spending-Augmented-Virtual-Reality-Forecast-Reach.

Yang, Jie, Yangsheng Xu, and Chiou S Chen. 1997. "Human Action Learning via Hidden Markov Model." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 27 (1): 34–44.