INTUITIVE BEHAVIOR

The Operation of Reinforcement Learning in Generative Design Processes

DASONG WANG[1] and ROLAND SNOOKS[2]
[1,2]*RMIT Architecture | Tectonic Formation Lab*
[1,2]*{dasong.wang|roland.snooks}@rmit.edu.au*

Abstract. The paper posits a novel approach for augmenting existing generative design processes to embed a greater level of design intention and create more sophisticated generative methodologies. The research presented in the paper is part of a speculative research project, Artificial Agency, that explores the operation of Machine Learning (ML) in generative design and robotic fabrication processes. By framing the inherent limitation of contemporary generative design approaches, the paper speculates on a heuristic approach that hybridizes a Reinforcement Learning based top-down evolutionary approach with bottom-up emergent generative processes. This approach is developed through a design experiment that establishes a topological field with intuitive global awareness of pavilion-scale design criteria. Theoretical strategies and technical details are demonstrated in the design experiment in regard to the translation of ML definitions within a generative design context as well as the encoding of design intentions. Critical reflections are offered in regard to the impacts, characteristics, and challenges towards the further development of the approach. The paper attempts to broaden the range and impact of Artificial Intelligence applications in the architectural discipline.

Keywords. Machine Learning; Generative Design Process; Multi-Agent Systems; Reinforcement Learning.

1. Introduction

The discipline of architecture is rapidly absorbing machine learning methodologies and searching for meaningful applications of these algorithms to design. This paper proposes an approach that augments known generative design strategies through a heuristic strategy that hybridizes a Reinforcement Learning (RL) based evolutionary approach with self-organizing generative processes. In contrast to the popular appropriation of image-based Generative Adversarial Networks that operate on the surface effects of architecture, this approach seeks to heuristically train generative algorithms that deeply embed architectural design intention within the algorithmic generative process.

Self-organizing algorithms that underlie complex systems have emerged over the past two decades as the basis for highly creative and volatile design strategies

PROJECTIONS, Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2021, Volume 1, 101-110.

that privilege bottom-up generative logic. Methodologies such as Behavioral Formation (Snooks, 2014), that engage multi-agent algorithms enable design intention to be embedded within the bottom-up rules of generative systems. However, these are often limited by their logic of local interaction that resists encoding systemic top-down design intention (Snooks, 2014).

This paper posits an approach in which the self-organizing behavioral logic of multi-agent algorithms is hybridized with the top-down evolutionary reward logic of Reinforcement Learning algorithms to enable the interaction of top-down and bottom-up design intention within a systemic design methodology. This augmented approach enables the hybridization of ideologically opposed, albeit complementary, algorithmic strategies to effectively compensate, or respond, to each algorithm's limitations.

The structure of the paper includes a background section covering the existing multi-agent and reinforcement learning algorithmic approaches, a methodology section that describes the hybridization of these approaches, and a discussion section that outlines the efficacy and implications of this hybridization. The posited approach is explored through a prototypical algorithmic framework that utilizes a mesh-based RL agent to evolve proto-architectural surface topology. While directly exploring the hybridization of multi-agent algorithms and reinforcement learning, this paper offers a generalizable approach to augmenting generative design.

2. Background

2.1. BEHAVIORAL FORMATION

Behavioral Formation is a generative design methodology that draws on the logic of swarm intelligence within a self-organized emergent process, which operates through Multi-Agent algorithms. This approach privileges encoded local interactions of autonomous computational agents, achieving intensive, intricate, heterogeneous phenomena in generative processes (Snooks, 2014). The exploration and development of this behavioral approach over the past two decades has expanded the realm of generative design paradigms (Leach, 2017).

Multi-Agent Systems (MAS), which trace their logic back to the Cellular Automata, are the key technical foundation of Behavioral Formation. The Multi-Agent algorithmic approach discussed here is an extension of the 'Boids' algorithm (Craig, 1987) which has been expanded and developed into various generative strategies including: 'Manifold Swarm' (Snooks, 2014), and 'Cellular Forms' (Lomas, 2014). Multi-Agent Systems depend on local interactions to generate self-organized emergence. However, this localized operation prohibits a more macro awareness and leads to an inherent limitation of global ignorance (Snooks, 2014). In practical design scenarios, this is regarded as a significant problem as a certain number of design intentions/constraints can only be defined from a global scale. For instance, the overall topology and structural performance of a form is a global issue. In order to augment the behavioral generative process, this paper purposes an approach that is based on the application of a complementary algorithm: Reinforcement Learning and Multi-Agent Systems.

2.2. ARTIFICIAL INTUITIONS

Artificial Intuitions speculate on a design process driven by Machine Learning (ML), with the capacity to develop typical and specific generative 'intuitions' towards quantitative subjective intentions and objective constraints of design (Wang, 2020). In regard to the scenario of Behavioral Formation, artificial intuitions are global awareness of encoded design intentions to be hybridized with the local interactions. The approach operates with the implementation of a Reinforcement Learning algorithm.

Reinforcement Learning algorithm is a long-standing Machine Learning framework with close association of optimal control. The mechanism of RL can be summarized as an agent seeks an optimal policy by interacting with its environment through feedback between observation states and quantified rewards, which can be traced back to a Markov Decision Process (Sutton, 1998). Different from other existing ML frameworks, RL is a heuristic approach that cultivates machine intelligence through training based on the accumulation of self-experiences instead of known human knowledge. The definition of RL is structured with several algorithmic elements: Agent, Observation States, Actions, Rewards and Policy (Henderson, 2018). Based on previous research (Wang, 2020), the paper demonstrates these elements through their application within a generative design context, as shown in Figure 1.

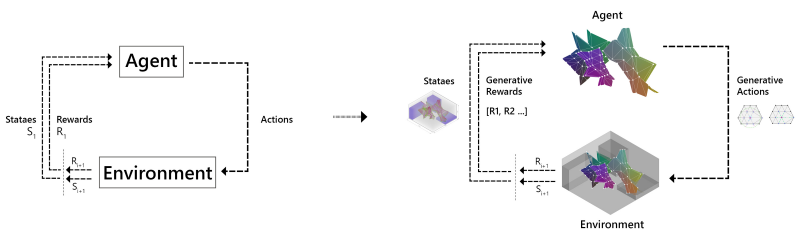

Figure 1. Translation of RL Algorithm within Generative Design Context.

- Agent: represents the entirety of the generative system itself (distinguished from each single agent in MAS). The Agent is capable of making decisions to undertake certain actions to achieve optimal generative outcomes.
- Action: demonstrates how the agent evolves and interacts with its environment through each step.
- Observation State: is a fixed-dimension matrix of information that summarizes the current status of the generative system and environment.
- Reward: assesses the performance of generative action with encoded quantitative criteria of design intention or constraints.
- Generative Policy: is the outcome of the training approach which can be regarded as the artificial intuition of agent. It maps states to actions, supervising the agent to take optimal decisions at each step to generate reward-oriented results.

2.3. COMPUTATIONAL FIELD

The term computational field within the paper refers to a strategy to represent three-dimensional topological information through a vector field. The mechanism of a computational field is to generate a 3D grid with a collection of vectors mapping to a series of information: position, direction, topological condition, etc. The strategy is implemented in the design process for two purposes: converting the agent's condition to a matrix data type for observation states, as well as bridging the generative outcome from training to behavioral formation process. This has been further discussed in Chapter 3.1 and 3.4.

3. Methodology

The design methodology of 'Intuitive Behavior' is demonstrated with a speculative design experiment aiming to explore and test the impact of the proposed RL approach by training a topological field with intuitive global awareness of pavilion-scale design criteria.

3.1. OVERALL METHOD

As shown in Figure 2, the overall design method is structured with three major steps:

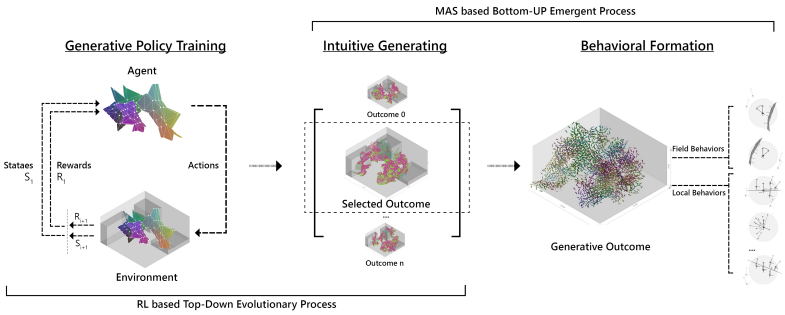

Figure 2. Overall Method of Structuring the RL based Intuitive Generative Approach.

- Generative-Policy Training: forms the main step added prior to the existing generative design procedure, which aims to train a generative system to achieve encoded global design intentions. The training process is inspired by and based on an evolutionary deep RL algorithm which is demonstrated in the following sub-chapters from 3.2 to 3.6 with particular emphasis.
- Intuitive Generating: is a process that a specific design context that will be inputted to generate potential design outcomes. An eventual generative outcome will be selected based on top-down unprogrammable criteria and forward to the next step.
- Behavioral Formation: The selected outcome will be used as a global topology for the Behavioral Formation process aiming to increase design resolution and

intricacy. The behaviors in the process will be encoded both in regard to local interactions and global response to the generated global topology.

3.2. RL AGENT AND ENVIRONMENT

Within the Generative-Policy Training process, the RL Agent is defined as a mesh graph that consists of certain vertices and predefined topology. As initialization of each training episode, the position and orientation of the mesh-Agent will be randomly generated as well as individual vertices (Figure.3). The purpose of using a mesh graph as RL Agent is due to its characteristic of efficiency, controllability, and flexibility in terms of representing a three-dimensional geometry. While the random initialization operation aims to expand the optimal solution pool with as much as potential possibilities.

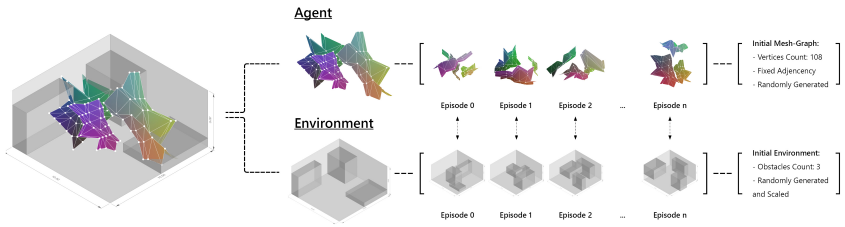

Figure 3. Strategies of Initializing RL Agent and Environment.

The environment is defined as a cuboid space with a scale of 10*10*5 meters, (Figure.3), in which several obstacles will be randomly generated as site conditions and constraints in order to train the adaptability of the generative policy with the capability of unpredictable initial conditions.

3.3. RL ACTIONS: STOCHASTIC + GENERATIVE BEHAVIORS

In the experiment, the agent action is encoded as a combination of stochastic RL actions and topological generative behaviors. Stochastic action is a typical strategy used in the RL algorithm that each vertex makes a random decision to move towards the six-axis. The stochastic actions will later be supervised by the algorithmic training process. Simultaneously, the vertices will also undertake topological behaviors: spatial separation and tension cohesion (Figure 4) which is encoded with the intention to maintain the topological condition of the geometry from unexpected intersections. The strategy of integrating stochastic RL actions with generative behaviors is inspired by the description of intelligent agency in ‘Complex Adaptive Systems’ (Holland, 1992), that each Agent is expected to make individual decisions but also subjects to interactive behaviors to form an emergent complex system.

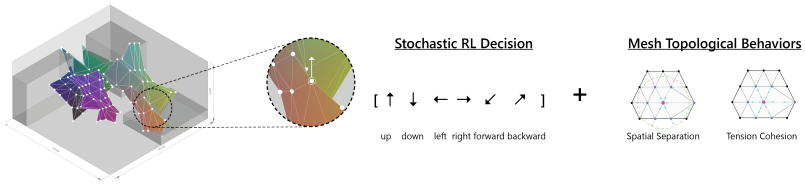

Figure 4. Strategies of Encoding RL Actions: Stochastic RL Decisions + Mesh Topological Behaviors.

3.4. RL OBSERVATIONS: COMPUTATIONAL FIELD

In response to the technical character of RL observation states, the RL Agent and Environment will be converted into a computational field with a three-dimensional voxel grid. Each voxel within the field contains informative states of the Agent and Environment (Figure 5). The states will then be calculated into a weighted value in order to structure a three-dimensional matrix for the Artificial Neural Network to proceed.

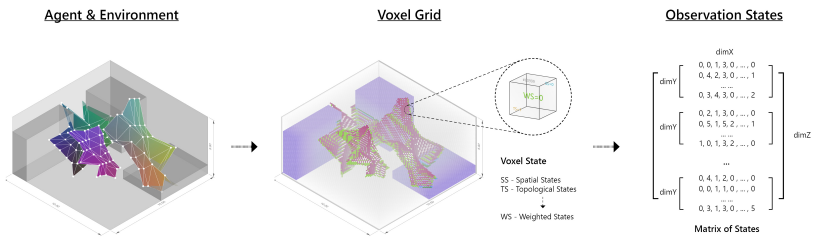

Figure 5. Strategies of Encoding Observation States.

3.5. RL REWARDS: GLOBAL DESIGN INTENTIONS

As the most critical element of the training process, the reward is defined towards global/macroscopical design intention that is not capable of a local behavioral system to achieve.

- Spatial Coverage (R1): is a criterion quantified based on a design intention of forming a mushroom-like topology. The design outcome is expected to generate spatial coverage above a certain height but remaining a minimum amount of coverage as a foundation (Figure 6).

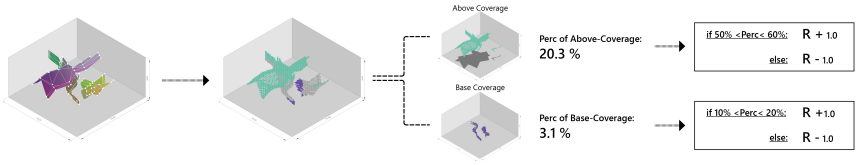

Figure 6. RL Reward 01 - Spatial Coverage.

- Topological Performance (R2): is defined based on the design intention of optimal structure-oriented topology. In the experiment, to avoid heavy calculation, a customized structure analysis algorithm is developed which calculates the topological states of each voxel based on the adjacent voxels below it (Figure 7).

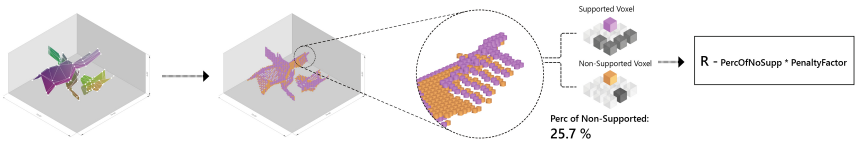

Figure 7. RL Reward 02 - Topological Performance.

- Site/Environmental Response (R3): is quantified from objective design constraints of collision avoidance with an unpredictable environment. As shown in Figure 8, the overlapping percentage of the form will be calculated each step and converted into a negative figure as a reward signal.

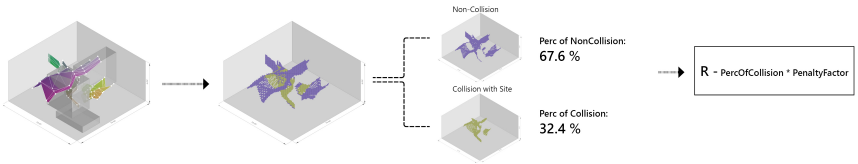

Figure 8. RL Reward 03 - Site/Environmental Response.

- Hierarchical Reward (HR): is defined to encourage the generative system (Agent) to achieve all the encoded rewards simultaneously. An additional reward signal will be added since the agent achieve more than any two rewards defined above, which increases exponentially according to the count of satisfied reward task.

3.6. TECHNICAL SETTINGS

The generative training process is based on a Deep Reinforcement Learning algorithmic framework, Proximal Policy Optimization (PPO), implemented on

the Unity platform with ML-Agent toolkit (Brockman, 2018). The technical settings and parameters of the experiment are listed in Table 1, with regard to the instructions from ML-Agent documentations.

Table 1. Parameters and Settings of Deep Reinforcement Learning Algorithm.

RL Hyper-Parameters	Network-Settings	Reward-Signals
batch_size: 128 buffer_size: 2480 learning_rage: 0.0003 beta: 0.005 epsilon: 0.2 lambd: 0.95	normalize: true hidden_units: 512 num_layers: 3 vis_encode_type: simple	extrinsic: gamma: 0.995 strength: 1.0

3.7. INTERACTIONS WITH BEHAVIORAL FORMATION PROCESS

As demonstrated in sub-chapter 3.1, an intuitively generative outcome will be selected and proceed as a global field for secondary behavioral formation processes. The paper speculates on several design strategies to demonstrate how two approaches being hybridized. Firstly, multiple agents are populated within the global field as initial states, which undertake both local behaviors and global interactive behaviors with the field to balance the local intricacy and global awareness. In Figure 9, a typical 'Manifold Swarm' (Snooks, 2014) generative strategy is implemented to continuously conduct the generative process.

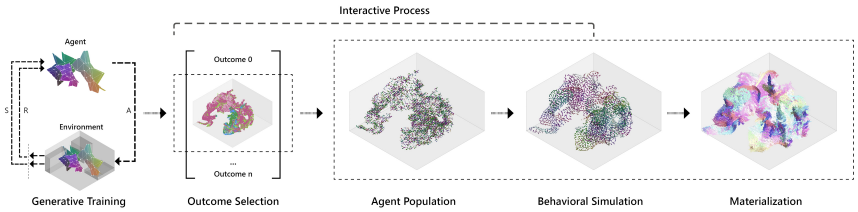

Figure 9. Interactions between RL Training Outcome and Behavioral Formation Process.

4. Discussions

4.1. GENERATIVE TRAINING OUTCOMES

The demonstrated RL-based generative training experiment has been conducted with a total episode of 20,000 (500 steps each). The generative training outcomes are recorded every 200 episodes, as well as the mean-reward recorded every 500/2000 episodes. As shown in Figure 10, the procedural generative outcomes present an obvious tendency of evolving from poor (blue) to optimal (red). This can be also observed from the mean reward chart that the performance of the

generative system increases rapidly in the first 5,000 episodes and gradually improves in the following with minor fluctuations. Based on the performative data, the paper speculates that the result can be more robust with additional training episodes in response to the complexity of the training scenario. Two procedural outcomes are highlighted to illustrate the intuitively evolving effect in response to the three predefined global rewards of spatial coverage, topological performance, and site/env response.

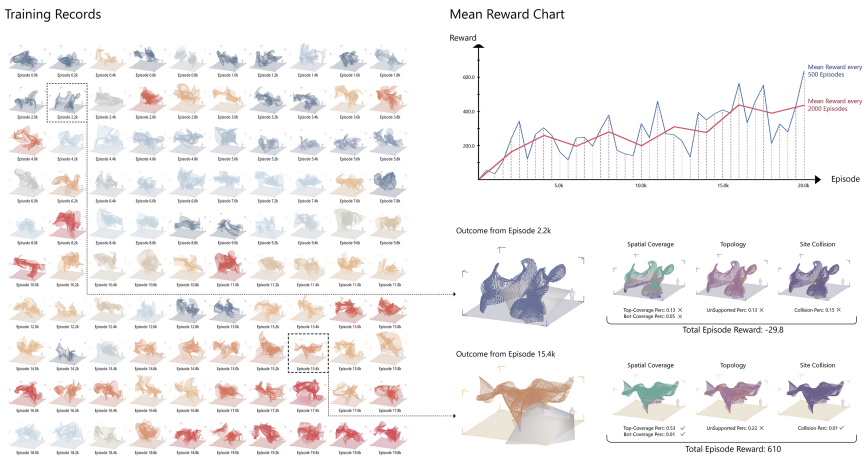

Figure 10. RL Training Processes and Outcomes.

4.2. REFLECTIONS ON THE APPROACH

Based on the training experiment, several reflections are summarized in regard to its characteristics, potential impacts, and challenges. Within the designing/training experiment, the generated global fields have evolved from a primordially stochastic status to be more intelligent and adaptive, forming an autonomous self-optimal process to potentially overcome the global ignorance of behavioral processes of formation. The Deep Reinforcement Learning approach demonstrates a heuristic procedure to computationally evolve generative policies for particular design criteria independent of direct human knowledge. This heuristic strategy enables the generative process to produce large quantities of outcomes that resist habitual aesthetics and formal tropes but still satisfy the predefined design intentions. Besides the application to behavioral processes of formation, the approach also shows significant potential to be applied in other generative processes, as its several key algorithmic components (Agent, Actions, Reward, etc.) can be fully customized in regard to various generative algorithms.

The experimental case described here hybridizes two independent approaches of intuitive training and behavioral formation, instead of directly training each single agent with global generative intentions. This is subject to the fact that

the RL algorithm is under development dealing with Multi-Agent circumstances. Emerging techniques have been proposed to tackle the problem, such as MADDPG (Multi-Agent Deep Deterministic Policy Gradient) algorithm (Lowe, 2020), which presents a promising methodology to train multi-agent algorithms with respect to global rewards. A series of technical challenges exist in terms of applying the proposed approach in a wider range. Firstly, the RL algorithm is an extremely computationally expensive process, fundamentally due to its heuristic character. In this case, the design intentions (rewards) are expected to be general rather than specific, which enables the training outcome to be adaptable in various analogous circumstances. Besides, the balance of precision and efficiency of the RL reward definition is also a challenge to achieve a reasonable result. As shown in the Intuitive Field experiment, the rewards definition is relatively lightweight instead of implementing a professional analysis algorithm (FEM structural assessment, for instance), which involves heavy calculations of million-magnitude iterations.

5. Conclusion

By hybridizing the intuitive field and behavioral formation processes, the RL based approach shows a meaningful impact on augmenting generative design with intuitive capacity and sophisticated control. Through the design experiment, the research presents a series of theoretical strategies and technical definitions for structuring a generative training process and evolving it through encoded design intentions. An augmented approach to the application of Machine Learning techniques has been proposed with the intention to broaden the range and impact of AI applications in the architectural discipline. The subsequent reflections in the paper also anticipate an interactive correlation between designers and computational intelligence.

References

Brockman, G., Cheung, V., Pettersson, L., Schulman, J., Tang, J. and Zaremba, W.: 2016, "OpenAI Gym" . Available from <arXiv:1606.01540> (accessed 12th December 2020).

Craig, R.: 1987, Flocks, herds and schools: A distributed behavioral model, *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*.

Henderson, P., Bachman, R., Precup, J. and Meger, D.: 2018, Deep Reinforcement Learning that Matters, *The Thirty-Second AAAI Conference on Artificial Intelligence(AAAI-18)*.

Holland, J.: 1992, *Complex Adaptive Systems*, The MIT Press.

N. Leach and R. Snooks (eds.): 2017, *Swarm Intelligence: Architectures of Multi-Agent Systems*, Tongji University Press, Shanghai.

Lomas, A.: 2014, Cellular Forms: an Artistic Exploration of Morphogenesis, *AISB-50*.

Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P. and Mordatch, I.: 2020, "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments" . Available from <arXiv:1706.02275> (accessed 12th December 2020).

Snooks, R.: 2014, *Behavioral Formation: Multi-Agent Algorithmic Design Strategies*, Ph.D. Thesis, RMIT.

Sutton, R. and Barto, A.: 1998, *Reinforcement Learning: An Introduction*, The MIT Press.

Wang, D. and Snooks, R.: 2020, Artificial Intuitions of Generative Design: An Approach based on Reinforcement Learning, *Proceedings of CDRF 2020*.