

High-Dimensional Evolution

Bringing Neural Networks and Evolutionary Simulations Onto Common Ground

Anna Gajdova, Tushant Jha, Erick Oduniyi, Lincoln Ombets, Edgar Valdes, Jana Weber,
*Artemy Kolchinsky
*Original proposal

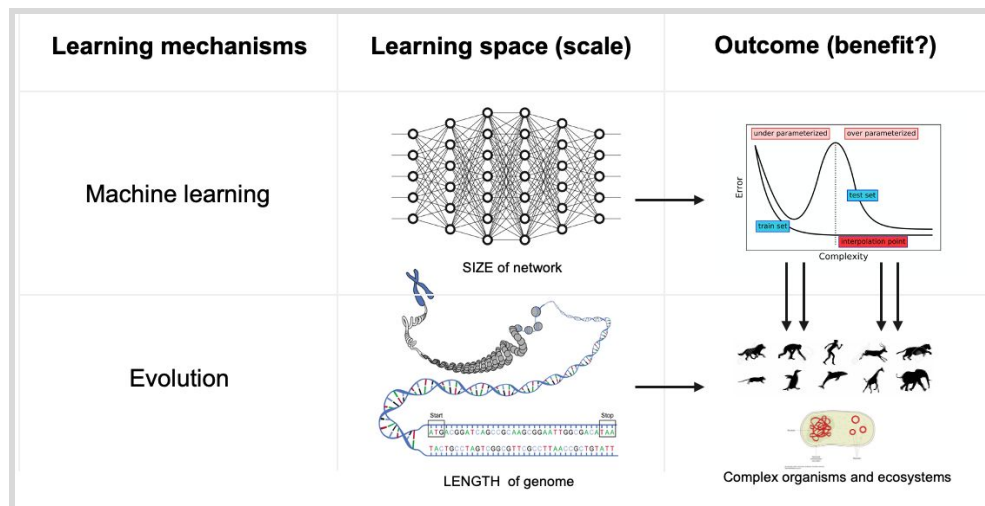


Figure 1: Intuitive parallels between ANNs and biological genomes, and key questions that motivate the present proposal.

Artificial neural networks (ANNs) are typically composed of interconnected nodes parameterized by large matrices of trainable parameters. Recently, generalization error has been found to decrease once the network size passes a certain point, in contrast to the traditional machine learning bias-variance tradeoff, which holds that increasing network size past a certain point will always lead to overfitting the training data. It is theorized that the massive number of parameters somehow "regularizes" the network performance [1] (and also see [2] for some theoretical work). Leading image classification networks now contain many millions or billions of parameters. This raises the question of how training occurs in such high-dimensional spaces.

In biology, all organisms have DNA-based genomes that encode information for the regulated synthesis of RNA and, ultimately, proteins that carry out most structural and functional roles. There is a rough correlation between genome length and organismal complexity, in the sense that single-celled organisms such as bacteria and yeast tend to have genome lengths on the smaller end – *a few million base pairs*; while multicellular organisms such as plants and animals tend to have genomes lengths on the longer end – *billions of base pairs* [5]. This observation inspires us to ask: Could there be an analogous benefit derived from a larger genome size in terms of the space of form and function that can be explored? (see **Figure 1.**)

To answer this and other questions, we propose constructing a unified framework for comparing ANNs and models of biological evolution [3, 4]. A central problem in doing so is the ill-defined nature of "evolutionary fitness." Therefore, we will develop a familiar artificial world in which agents guided by ANNs and agents guided by models of evolving organisms can face the same problem-solving or exploration tasks. The study's primary objects will be the agent-agent, agent-environment, and environment-environment interactions that emerge under diverse models of (1) agent behavior and capability; (2) environment structure and spatiotemporal variation; and (3) optimization or fitness. Further, the adapting agents' high-dimensional trajectories should provide novel insights into both ANN training and biological evolution, as shown in **Figure 2**.

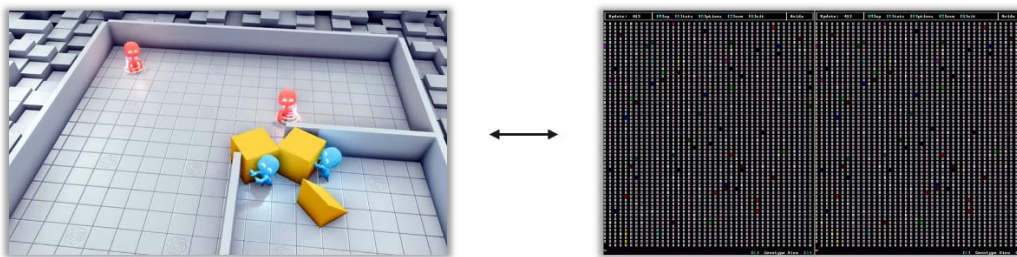


Figure 2: Example of Neural Agents (OpenAI Gym) vs. Evolutionary Populations (Avida 2.6). The group's work aims to create a framework (and subsequent application) where model comparisons and transfer learning between is supported through joint/mutual agent simulation.

References

- [1] Belkin et al., Reconciling modern machine-learning practice and the classical bias–variance trade-off, PNAS, 2019.
- [2] Advani and Saxe, High-dimensional dynamics of generalization error in neural networks, arXiv:1710.03667.
- [3] Gavrilets, Evolution and speciation on holey adaptive landscapes, Trends in Ecology & Evolution, 1997.
- [4] Conrad, The geometry of evolution, BioSystems, 1990.
- [5] Milo and Phillips, *Cell Biology by the Numbers*. “How big are genomes?” URL: <http://book.bionumbers.org/how-big-are-genomes/>