Generative Decision Making Under Uncertainty

Aditya Grover

University of California, Los Angeles
adityag@cs.ucla.edu

The ability to harness vast amounts of labeled and unlabeled data for efficient and accurate inference is a long-standing goal of artificial intelligence (AI). In recent years, several machine learning (ML) models for natural language processing and computer vision have successfully leveraged diverse datasets in a two-step paradigm: (a) pretrain deep neural networks on large unlabeled datasets e.g., text and images on the Internet; (b) adapt these models to any target domain using finetuning or few-shot prompting. This paradigm generalizes exceedingly well, achieving state-of-the-art performance on numerous tasks in language and perception.

While these results are encouraging, the capabilities of natural agents go much further in learning to *interact* with their physical environments without excessive supervision. For example, human infants develop strong intuition about the dynamics and control of physical objects by playing with block toys and fusing audio, visual, and haptic feedback, without any adult supervision. How can AI agents learn to flexibly interact using in the wild datasets? Specifically, we will focus on data-driven approaches for two paradigms in interactive decision making: reinforcement learning (RL), where an agent learns to make decisions via trial-and-error; and black-box optimization (BBO), where an agent learns to optimize an unknown function using pointwise evaluations.

First, we discuss the pretraining-finetuning regime for reinforcement learning. A conventional approach for offline pretraining simply adapts a *forward* RL algorithm, such as Q-learning or model-predictive control, into an offline algorithm that can work with large off-policy datasets. However, extracting policies from such forward models suffers from distribution shifts. In contrast, we focus on *inverse* models, which directly map outcomes (e.g., returns) to actions conditioned on the state of the environment. Since these inverse mappings are typically one-to-many, we will discuss the use of deep generative models for learning return-conditioned policies. One such instantiation, Decision Transformers (DT), learns an autoregressive generative model that parameterizes the agent policy using a Transformer architecture (Chen et al. 2021). We show DT is stable to train using a vanilla supervised objective and highly effective at modeling long-range dependencies, especially in environments with a sparse reward structure. Further, unlike forward models, DT is reliable at generalizing to out-of-distribution returns (Nguyen, Zheng, and Grover 2022). Above and beyond offline pretraining, we demonstrate excellent finetuning of such models using a novel entropy-based regularization scheme and hindsight relabelling (Zheng, Zhang, and Grover 2022).

Second and last, we will discuss generative models for black-box optimization. Similar to the RL case, standard approaches are based on learning forward surrogates for the black-box function that generalize poorly outside the offline dataset. Inverse approaches, powered by powerful generative models such as autoregressive transformers and diffusion models (Mashkaria, Krishnamoorthy, and Grover 2023; Krishnamoorthy, Mashkaria, and Grover 2023), can successfully generalize to function values outside the offline dataset using novel forms of attention and guidance. Empirically, we demonstrate state-of-the-art performance on design benchmarks drawn from diverse domains in sustainability, biology, and engineering. We will conclude with some open directions for future investigation, including the use of generative models for multi-tasking in complex scenarios involving multiple task specifications (Liu et al. 2022), competing objectives (Zhu, Dang, and Grover 2023), and exploration constraints (Du, Abbeel, and Grover 2022).

References

Chen, L.; Lu, K.; Rajeswaran, A.; Lee, K.; Grover, A.; Laskin, M.; Abbeel, P.; Srinivas, A.; and Mordatch, I. 2021. Decision transformer: Reinforcement learning via sequence modeling. *NeurIPS*.

Du, Y.; Abbeel, P.; and Grover, A. 2022. It Takes Four to Tango: Multiagent Selfplay for Automatic Curriculum Generation. *ICLR*.

Krishnamoorthy, S.; Mashkaria, S. M.; and Grover, A. 2023. Diffusion Models for Black-Box Optimization. In *ICML*.

Liu, F.; Liu, H.; Grover, A.; and Abbeel, P. 2022. Masked Autoencoding for Scalable and Generalizable Decision Making. *NeurIPS*.

Mashkaria, S. M.; Krishnamoorthy, S.; and Grover, A. 2023. Generative Pretraining for Black-Box Optimization. In *ICML*.

Nguyen, T.; Zheng, Q.; and Grover, A. 2022. Reliable Conditioning of Behavioral Cloning for Offline Reinforcement Learning. *arXiv preprint arXiv:2210.05158*.

Zheng, Q.; Zhang, A.; and Grover, A. 2022. Online decision transformer. *ICML*.

Zhu, B.; Dang, M.; and Grover, A. 2023. Scaling Pareto-Efficient Decision Making via Offline Multi-Objective RL. In *ICLR*.