Week 10 Paper Review

Niharika Shrivastava
School of Computing
National University of Singapore
Singapore, 119077
niharika@comp.nus.edu.sg

Abstract

This is a brief review of [1], [2], [3].

1 Deep Visual Foresight for Planning Robot Motion

This paper demonstrates an approach for robotic tasks that require minimum human involvement or supervision and can learn without a detailed reward function, an image of the goal, or ground truth object pose information by combining deep action-conditioned video prediction models with MPC.

Primarily, the goal is defined as moving a set of pixels from their initial positions to the target position which aids arbitrary rearrangement of objects. Specifically, the implementation heavily uses a deep video prediction LSTM model (previous work) that predicts flow operators. These are used (instead of predicted images) to transform prior pixel distributions into pixel distributions for the next state, thereby representing independent states using which the highest probability actions can be inferred in a model-based setting.

Experiments show that the method formulation surpasses the problem of high-dimensional input by representing state transitions in the form of pixel value changes. The baseline comparisons are somewhat naive but the proposed approach does show potentially significant results in comparison. However, although it is general to completely novel objects, the model is still limited to relatively simple short-horizon tasks. Overall, given the limited input required to perform tasks, it is a promising approach that will become better with increased performance in video prediction models.

2 Exploring Model-based Planning with Policy Networks

The authors propose POPLIN, an MBRL algorithm that formulates action planning at each time step as an optimization problem using neural networks. The main idea is to optimize the policy network's parameter space for greater efficiency and utilization of deep neural networks as they seldom get stuck in sub-optimal points.

Specifically, the implementation builds on PETs (which uses a probabilistic ensemble to capture model uncertainty) by adding noise in the parameter space of the policy network, enabling it to be sample efficient and have better asymptotic performance even for high dimensional environments. Furthermore, distilling knowledge from planning trajectories into a policy network after interacting with the environment to altogether avoid the expensive online planning is also explored, which achieved high performance on Cheetah.

Several ablations and varied initialization methods are discussed for either optimizing in action space or parameter space. POPLIN-P (optimization in parameter space) performed better for large population sizes than POPLIN-A (optimization in action space). Moreover, it was shown that policy control is not always successful compared to MPC control owing to distillation collapse in deterministic policy networks.

Preprint. Under review.

3 Planning with Goal-Conditioned Policies

The authors propose the Latent Embedding for Abstracted Planning (LEAP) framework that combines model-free RL that provides temporal abstractions to learn low-level goal-conditioned policies with model-based planning over a latent variable representation of sub-goals. Primarily, two-level abstractions are made: temporal abstractions allow planning at a coarser time frame, while state abstractions allow planning over simpler state representations. This motivates to solve tasks with high-dimensional inputs like images and benefit from the compositional structure of planning. It builds on TDMs that help to determine the reachability of sub-goals generated using the state abstractions and executes actions predicted using a goal-conditioned policy to reach these goals. The sub-goals themselves are a learned latent representation using VAEs.

LEAP not only performed significantly better than prior works such as PETs, TDM, etc. on image-based robot navigation and manipulation tasks but also long-horizon complex non-vision-based tasks such as Ant Navigation. However, they do not address the possibility of exploration for goal-conditioned policies and leave it to future work.

References

[1] Finn, C., & Levine, S. (2016). Deep Visual Foresight for Planning Robot Motion. ArXiv. https://doi.org/10.48550/arXiv.1610.00696

[2] Wang, T., & Ba, J. (2019). Exploring Model-based Planning with Policy Networks. ArXiv. https://doi.org/10.48550/arXiv.1906.08649

[3] Nasiriany, S., Pong, V. H., Lin, S., & Levine, S. (2019). Planning with Goal-Conditioned Policies. ArXiv. https://doi.org/10.48550/arXiv.1911.08453