Model Predictive Curiosity for Self-Supervised Dynamics Models

Motivation

→ A crucial element of child psychological development is *play* (with objects like blocks, toys, etc.) - learn about the physical dynamics of objects and environments around them through interaction and observation.

→ Developing auxiliary reward that encourages unconstrained exploration is the computational analog of this process

- Agent explores an environment and takes actions that are not conditioned on any particular end-goal.

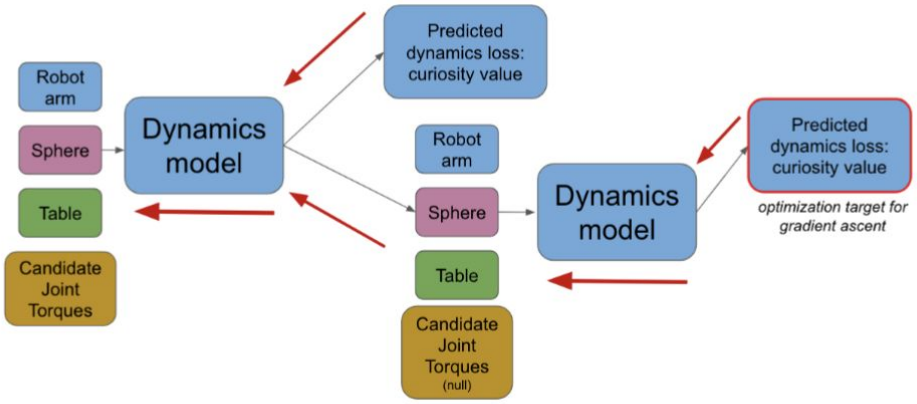
→ Key Question: How does active exploration and self-supervised learning facilitate the learning process of reinforcement learning agents?

- Modeling process of active exploration and self-directed learning, so as to develop AI agents capable of learning flexibly and robustly in a similar way.

Model Predictive Curiosity (MPCu)

→ A framework inspired by the Model Predictive Control (MPC) paradigm which maps state-action pairs to curiosity values, thus predicting the error in the forward prediction model.

→ Model backpropagates from the predicted curiosity value to select the action which would maximize this value.



Related Works

1. Towards Curiosity-Driven Learning of Physical Dynamics
2. Deep Visual Foresight (Finn, 2016)
3. Visual Interaction Networks (Watters et al.)

Stanford University, PSYCH 240A
Michael Lingelbach, Olivia Lee, Priscilla Zhao

Model Predictive Curiosity for Self-Supervised Dynamics Models

Methods

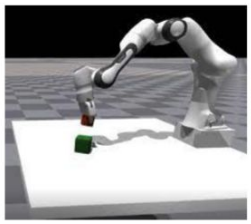
→ Goal implementation in NVIDIA's IsaacGym physical simulation environment (Fig. 1a), Box2D used for testing. Model uses *object-centric embeddings* to represent arm/environment (Fig. 1b)

→ Dynamics model uses sequence-to-sequence forward prediction (Fig. 2), incorporated into MPCu for sequence-to-sequence-based curiosity estimation

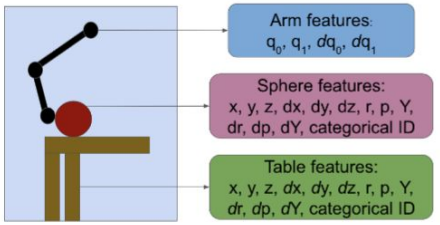
- Samples several rollouts from dynamics model, to compute expected loss over multiple timesteps.
- Maximize predicted loss in dynamics model using curiosity value as optimization target for gradient ascent.
- Like MPC, predict future world state at time *t+1* and predicted loss (L2 pose estimation error).
- From prediction target, perform gradient ascent via backpropagation through dynamics model into candidate joint torques → choose joint torques which maximally antagonize dynamics model.

→ Can directly leverage dynamics model for control by performing traditional MPC

- Ignore forward pass through curiosity model
- Use the desired world state as the optimization target for gradient ascent, to minimize the delta between desired and predicted world state.



(a) Sample IsaacGym environment with robotic arm, tabletop, and soft block.



(b) Simplified environment: 2-DoF robotic arm, tabletop, and sphere

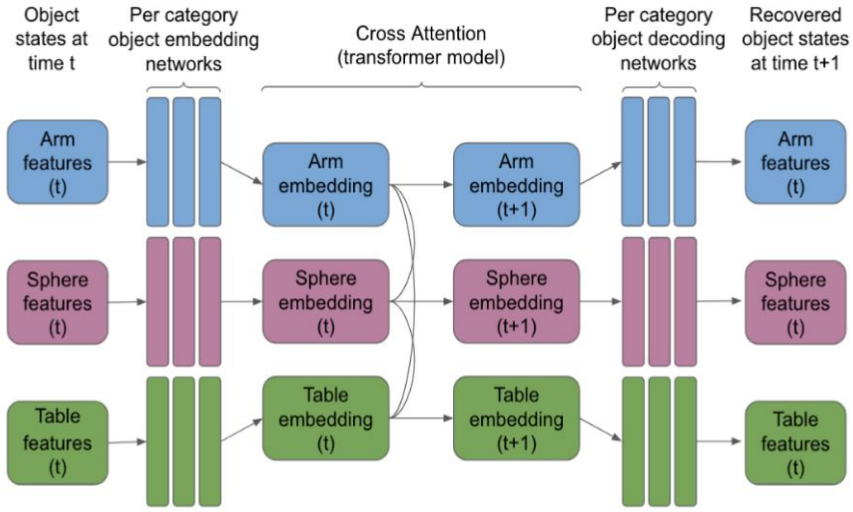


Figure 2: Proposed dynamics model using sequence-to-sequence based forward prediction on object-centric embeddings

Stanford University, PSYCH 240A
Michael Lingelbach, Olivia Lee, Priscilla Zhao

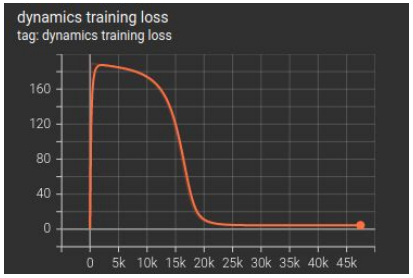
Model Predictive Curiosity for Self-Supervised Dynamics Models

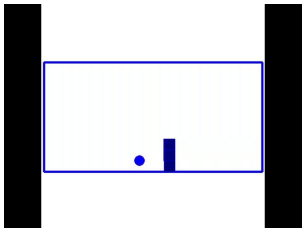
Results

→ Generated 50000 training scenarios of a force being applied to a circle adjacent to a tower in a Box2D environment

→ Trained a dynamics model to predict forward motion of circle, and a curiosity model to predict the loss in the dynamics model

→ At test time, used a backward pass through the curiosity model into the action space to perform gradient ascent on the force vector (not origin of the force, which was fixed to the circle)





Pre-MPCu optimization of action trajectory

Post-MPCu optimization of action trajectory

Discussion

→ MPCu is capable of directly optimizing for high curiosity action values

→ In the simple environment, MPCu enriches forces that cause multi-object interactions

→ MPCu depends on the predictive power of the model, accuracy of the curiosity prediction network, and the inductive bias for action selection (initialize action on the objects)

Conclusion & Future Work

→ Performance of system hindered by dynamics model choice, more work is needed to develop a good dynamics model for MPCu

→ Extend experiments to IsaacGym environment and embodied Box2D environment

→ Incorporate longer horizon rollouts into action selection phase

Stanford University, PSYCH 240A
Michael Lingelbach, Olivia Lee, Priscilla Zhao