Review-14

Skill Transfer Via Partially Amortized Hierarchical Planning

Solving new tasks in complex environments requires effective transfer of knowledge. To this end, the work proposes a Learning Skills for Planning (LSP) for transfering skills in a partially amortized framework by adopting hierarchical planning for selection of skills. A Reinforcement Learning (RL) agent plans to compose skills in imagination which are used to condition the low-level behavioral policy for learning and optimally acting in the environment. Planning is carried out be making use of a world model which allows sample-efficient learning. Planning in low-dimensional skill-space allows the agent to suitably transfer knowledge across different tasks.

Hierarchical planning in a partially amortized setting combines the fully online and fully amortized methods for acting. Online planning is carried for selecting skills which are used to condition on the low-level policy. Agent optimizes its low-level policy in the fully amortized setting. Learning of the agent is carried out by making use of a world model similar to Dreamer. Model learning consists of training the representation, observation, latent-state and task-reward modules. High-level skills are held fixed for K number of steps and optimized using CEM. Behavioral learning consists of training the lowe-level policy in world model using the skills distribution and backward skill predictor. Lastly, the policy interacts in the environment using MPC with a pre-sampled high-level skill. In order to facilitate learning of skills, the method constructs an auxilary objective which maximizes the Mutual Information (MI) between latent skills and state sequences. This requires a tractable distribution which is adopted using the backward skill predictor and trained using supervised learning in conjunction with reparameterization of CEM distribution.

Hierarchical planning of skills facilitates sample-efficient transfer of knowledge on complex locomotion tasks. Necessity of planning in skill-space is highlighted by comparing the performance of proposed method in the absence of skill planning. However, the method presents several shortcomings. Firstly, the world model is fairly similar to Dreamer and based on its latent state-space model. Upon comparing LSP with Dreamer on MI ablations and knowledge-transfer tasks, one can observe that LSP demonstrates high variance and is only comparable with Dreamer. This indicates the limited necessity and effectiveness of skill planning since the same performance can be achieved with Dreamer on all tasks. Secondly, LSP reports its performance only for 3 random seeds. This may not be sufficient for validating comparison to its baselines and confidently accounting for stochasticity in the environment. Lastly, LSP presents its motivation to avoid planning in high-dimensional action spaces and effectively plan skills in the low-dimensional skill space. However, experiments carried out on locomotion and transfer tasks do not demonstrate this insight. For instance, one could compare planning of LSP and its baseline algorithms on tasks with high-dimensional action spaces such as the Humanoid.