11/20/2025

Parameter-Space Policy Composition for Sim-to-Real Transfer in Quadruped Locomotion Control

This presentation addressed the long-standing challenge of sim-to-real transfer in robotic locomotion, focusing on quadruped robots operating under varying environmental conditions. While reinforcement learning (RL) enables robots to acquire adaptive behaviors through interaction, training directly in the real world is costly and impractical due to factors such as hardware wear, data inefficiency, and safety constraints. As a result, most modern approaches rely on training policies in simulation and deploying them on real robots. However, discrepancies between simulated and real dynamics often lead to performance degradation.

The dominant strategy to mitigate this gap is domain randomization, where environment parameters such as mass, friction, and initial states are randomly varied during training. Although this produces robust policies, the resulting behaviors are often overly conservative, as the policy must handle all possible variations simultaneously. Domain adaptation methods attempt to identify the real environment's dynamics more precisely, but they typically rely on a single large neural network trained across multiple environments, which limits scalability.

To overcome these limitations, the authors proposed a novel framework based on parameter-space policy composition. Instead of training a single monolithic policy, the method trains multiple expert policies, each specialized for a particular environment. A new policy for a target environment is then constructed by linearly combining these experts directly in parameter space using a set of base weights. Crucially, only these base weights need to be adapted when transferring to a new environment, enabling efficient adaptation with very few real-world trials.

The approach was inspired by prior work on policy composability and was implemented using a shared network structure with task-specific mixing coefficients. Training alternated between optimizing shared base parameters and environment-specific weights, ensuring that the resulting policies were suitable for linear combination.

Experiments were conducted on a quadruped robot performing a turning task under varying friction conditions. In simulation, four expert policies were trained using Soft Actor-Critic, each corresponding to a different friction level. The authors demonstrated that optimal turning behavior changes significantly with friction, validating the need for adaptive control. For unseen target environments, effective composite policies were obtained by searching for optimal mixing weights. While grid search was used in simulation, real-world experiments employed Bayesian optimization to minimize the number of required trials.

Real-robot experiments on low-, medium-, and high-friction surfaces showed clear performance improvements as base weights were optimized. The robot progressed from unstable behavior to fast and efficient turning within a limited number of trials, outperforming standard domain-randomized policies.