

Free Challenge for Simulation 3D-FC Portugal 3D Team

Contextual Policy Search for Learning a Flexible Bipedal Robot Locomotion Controller

Abbas Abdolmaleki, Nuno Lau, Luis Paulo Reis, Nima Shafii, Lucas Lemos, Rui Ferreira, Jan Peters¹, Gerhard Neumann²

Abstract—Black box optimization methods such as the evolutionary strategies, e.g. CMA-ES, have been employed extensively for gait learning, however these algorithms fail to generalize a learned gait to different contexts. In order to generalize the gait to, for example, different walking speeds, typically the parameters are learned for several target speeds independently. Subsequently, Regression methods are used to generalize the learned gaits to a new, unseen speed. Although such approaches have been used successfully, they are inefficient in terms of the number of needed training samples as the learning of the policy parameters for different contexts and the generalization between learned policies for different contexts, are two independent processes. Hence, we cannot reuse data-points obtained from learning of a gait for one context to improve and accelerate the learning gait of another context. Given these limitations, We investigated the application of contextual policy search to learn and generalize the learned parameters of a bipedal locomotion controller for multiple unseen contexts. To achieve this desired flexibility of the learned walking controller, we use a recent developed contextual policy search method called contextual relative entropy policy search (REPS). In our experiments we choose speed as the context.

I. THE APPROACH

We want to learn an upper-level walking policy $\pi(\theta|s)$ that sets the parameters θ of the lower level walking policy given a desired speed s . We model the lower level policy as a ZMP based walking controller that models the dynamics of the robot as an inverted pendulum model augmented with a spring model which has 10 parameters to set. Table 1 shows the walking parameters θ . To learn the upper level policy we use contextual REPS. Contextual REPS is an information theoretic policy search method. The main insight of using information theory is to bound the relative entropy, also called Kullback-Leibler divergence, between two subsequent policies. Please refer to the paper [1] for more details on REPS.

II. RESULTS

We perform a simulation experiment using a simulated NAO humanoid robot. The task is to learn a policy for walking straight for speed range of [0.1 0.8] as context s , starting from a complete

¹Jan Peters is with Intelligent Autonomous Systems in TU Darmstadt and Max Planck Institute for Intelligent Systems

²Gerhard Neumann is with Intelligent Autonomous Systems in TU Darmstadt

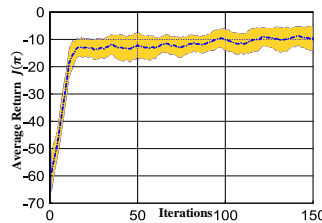


Fig. 1. Learning curve of learning the upper-level policy

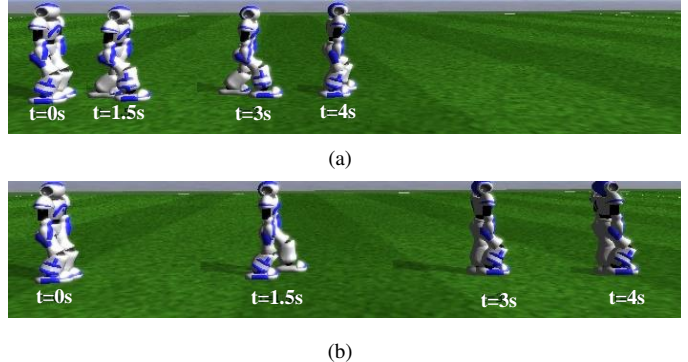


Fig. 2. (a) snap shot of learned gait for speed 0.38 m/s (b) snap shot of learned gait for speed 0.76 m/s

TABLE I

THE OPEN PARAMETERS OF LOWER LEVEL WALKING POLICY

Notation	Description
A	Oscillation Amplitude of the Spring
x_0	Offset Term of the Spring
β	Regulator of the Position of CoM Between The Feet
z_{step}	Maximum Height of The Swing Foot From The Ground
T	Duration of The Step
r_{offset}	Torso Roll Offset
γ_{tilt}	Feedback Gain For The Torso Tilt
γ_{roll}	Feedback Gain For The Torso Roll
v_x	The Step Size Along The x axis
ϕ_{DSP}	Duration of Double Support Phase

step. We define a speed dependent reward function as follows:

$$R(s, \theta) = \alpha(\dot{x} - s)^2 + \beta(\dot{y})^2 + \sigma(1_{fall})$$

where we choose $\alpha = 10$, $\beta = 1$ and $\sigma = 100$. Figure 1 shows the learning curve for learning the upper-level policy and Figure 2 shows two learned gaits for two different speeds.

III. CONCLUSION

We efficiently optimised and generalized the learned parameters of our walk controller for a continuous range of speeds which is demanding for humanoid locomotion. In the future we will test our method using various contexts such as different surface slopes and different surface frictions.

REFERENCES

- [1] Kupcsik, A.G.; Deisenroth, M.P.; Peters, J.; Neumann, G. (2013). Data-Efficient Generalization of Robot Skills with Contextual Policy Search, Proceedings of the National Conference on Artificial Intelligence (AAAI).