Stable Locomotion in Unstructured Terrain using Curriculum Learning for Online Parameter Adaptation

Sam Shaw, Mateo Guaman

1 Project Overview

Legged animals use sensory feedback to inhibit or extenuate certain gait characteristics online, allowing them to subconsciously navigate extreme terrain with ease [2, 8]. Inspired by nature, legged robots can similarly traverse complex environments inaccessible to other systems, such as wheeled robots. However, two major challenges are presented: 1) the coordination of the many degrees-of-freedom (DOF) often present on legged systems to produce locomotion, and 2) determining suitable higher-level gait parameters for a given environment. In the context of robotics, central pattern generators [3] (CPG) aid in solving the first problem. Often modelled as a set of coupled oscillators [7], CPGs allow for the generation of parameterized gaits and provide smooth transitions when gait parameters are changed. The limit cycle of the CPG determines the step shape, and other numerical parameters control specific aspects (e.g., step length, step height, step speed).

When the path of the robot and surrounding environment is known *a priori*, it is sometimes possible to set static parameter values (e.g., set a step height larger than the highest obstacle) or encode simple rules [9]. First, such methods are not suitable for general robot deployment, if the environment is not well-known. Second, such methods are purely reactive; a robot may locomote to a position in which it cannot appropriately react. Finally, these methods encode a significant amount of domain knowledge – which may or may not be optimal.

We are interested in ensuring stable, forward locomotion of a legged system on a variety of terrains – from flat ground to a flight of stairs. We will employ a learning-based approach to adapt gait parameters online based on sensory feedback. Specifically, we aim to use IMU feedback to determine 1) the slope of the terrain in the direction of locomotion and 2) stability of locomotion (by looking at changes in IMU feedback). By adapting gait characteristics such as step speed, step length, and center of mass position based on this feedback, we will enable a legged system to more naturally navigate obstacles and ensure stable forward locomotion. We aim to apply a curriculum learning approach to facilitate learning – increasing the complexity of the locomotive challenge as learning progresses. Specifically, the terrains in our curriculum will have a variety of obstacles and slopes. We aim to develop and evaluate our curriculum by performing a number experiments in a Gazebo simulation environment.

2 Background and Related Work

In this section, we present background information on curriculum learning before detailing previous works on learning-based, full-robot control.

2.1 Transfer and Curriculum Learning

In reinforcement learning problems, we aim to reduce the amount of resources used in the learning process while maximizing the reward that the agent obtains. For complex tasks, transfer learning can be used to reduce the resource requirements for a learning agent. The idea behind transfer learning [11] is that what has been learned from experience for one task can be reused to learn another, but different, task. Transfer learning has been successful in reducing learning time and increasing reward achieved for complex tasks [12] compared to learning from scratch.

We are concerned with Curriculum Learning, a flavor of transfer learning where an agent learns from examples presented to it in a particular order, usually in order of increasing complexity. It has been shown that curriculum learning provides significant improvements in machine learning problems in terms of better generalization, faster convergence speed, and at finding better local minima [1]. In reinforcement learning, curriculum learning can be used to develop sub-tasks as components of a multi-step curriculum with the goal of obtaining better performance in a complex task [5].

2.2 Full-robot Control via RL

The problem of using reinforcement learning to achieve full-body control of a robot has been explored in the past, using a variety of learning techniques. In [4], Kohl and Stone achieved full-robot control of a quadrupedal robot using policy gradient reinforcement learning. Their agent learned to search for gait parameters for locomotion on level, planar terrain that maximize the walking speed of the robot. Their approach significantly outperforms a variety of existing hand-coded and learned solutions. Reinforcement learning has also been used to achieve forward locomotion by learning motion patterns for different segments of a caterpillar robot [13].

Peng et al. [6] used Deep Reinforcement Learning to achieve full-robot control for terrain-adaptive locomotion. They used a mixture of actor-critic experts to achieve locomotion in multiple planar bio-inspired characters and terrain classes: gaps, steps, walls, and slopes. Specifically, their approach directly learns a policy for assigning joint angles and forces to simulated actors.

Sartoretti et al. [10] use an aynchronous advantage actor-critic (A3C) algorithm to learn a decentralized control approach for the locomotion of a modular, articulated snake. Specifically, using torque feedback, they to learn decentralized control policies that adapt shape parameters for different windows (segments of the snake). Their learned controller greatly outperforms a shape-based compliant control approach for forward progression in confined environments.

3 Problem Formulation and Technical Approach

We first define the learning problem precisely and present our approach before detailing our experimental validation procedure.

3.1 Learning Problem

In this project, we aim to learn a policy for adapting specific high-level gait parameters online based on IMU feedback for the stable, forward locomotion of a legged robot on unstructured terrain. In the context of this project, we intend to update desired step-length, step-speed, and center-of-mass position (along one dimension – in the direction of locomotion), considering updates to these parameters as discrete positive, negative, or zero increments. Thus, as stated above, we have twenty-seven possible actions: three possible gait parameters to adapt with three possible discrete increments to apply.

Formulating the learning problem will have two major challenges: 1) selecting features for state-space representation, and 2) designing an appropriate reward structure. As a first approach, we intend to define the current state such that it encompasses both the current values of the adapted gait parameters and additional feedback from the IMU. Specifically, we intend to define the state with the following features: step-speed, step-length, and center-of-mass position for locomotion; and body-roll angle and body-pitch angle for stability. Additionally, we anticipate that a suitable reward structure will involve providing the agent positive reward for forward motion and a negative reward for lack of body stability (this may involve taking smaller steps when the terrain is rougher).

3.2 Learning Approach

We cannot determine the stability of the locomotion based on the IMU feedback at any single timestep, since a large roll or pitch body angle could be observed during locomotion on a slope (even when the robot is stable). Instead, we can determine the stability of the robot by examining the change in IMU feedback over several iterations; gradual changes over time are indicative of terrain change, whereas rapid changes may indicate foot slippage or tipping. To capture the temporal changes in IMU feedback, we intend to use a Deep Recurrent Q-Network to learn a state-action mapping described above.

To facilitate learning, we intend to develop a learning curriculum. We will begin by training on planar, level ground; here we expect the agent to learn how to adapt its step length and speed to achieve reasonable stability. Next, we can begin to train on planar slopes; here we expect the agent to learn how to adapt its center-of-mass position on top of regulating step length and step speed. Finally, we can attempt trials on unstructured terrain – the ultimate goal. By first presenting more simple locomotive challenges, we aim to direct the agent to discover what types of environments (defined by IMU feedback) should induce what parameter change. We believe that partitioning the problem in this manner will allow for increased learning speeds.

3.3 Experimental Setup

For this project, we will perform experiments in a Gazebo simulation environment. The legged system that we will use for experimental validation is an 18-DOF hexapod robot. Each of the robot's 18 DOFs has encoders for position and torque measurements as well as a 3-axis gyroscope and 3-axis accelerometer for orientation in the world.

We will use ROS (Robot Operating System) as the communication infrastructure for the system. The learning agent communicates the learned gait parameters to a mid-level controller, which in turn translates the gait parameters to joint angles for the robot. These angles are then published to the Gazebo simulation, which directly uses them to simulate the movement of the robot. The new pose of the robot is then obtained from the Gazebo simulation and passed as feedback back to the learning agent.

In order to develop and validate a curriculum for the agent, we will generate different scenarios in Gazebo with increasingly levels of complexity in terms of obstacles and terrain. Learning in a simulation environment allows for faster learning, and avoids wear and tear of the hardware.

4 Evaluation and Expected Outcomes

To evaluate the performance of our gait-parameter control policy learned via a curriculum, we will perform two other types of simulated experiments: 1) learning directly from scratch on the hardest environment, and 2) with static gait parameters (no learning involved). This project will be successful if a learned policy out performs static parameter choices, and especially successful if a curriculum learning approach helps significantly reduce learning times or leads to a more-effective policy. We will determine performance using two metrics; we are concerned with both the learning speed, which is how quickly we converge to a policy, and the effectiveness of the policy, which will be measured based on cumulative reward.

We expect that the curriculum learning approach will greatly increase learning speed. We believe that slowly increasing the difficulty of locomotive challenges will guide the agent to learn how to adapt gait parameters one-by-one for both stability and forward motion. With regard to the effectiveness of the learned policy, we expect that for a fixed number of episodes, the policy learned by the curriculum approach will outperform a policy learned directly on steep, unstructured terrain.

5 Extensions

Future experiments could implement the final simulated approach using the robot's on-board IMUs. Additionally, future experiments could augment onboard IMU sensing (e.g., with lidar) with additional feedback on the environment's structure, necessary for anticipatory control. Doing so would allow the robot to both react to its environment, and also anticipate future parameter changes (e.g., an increase in step height to navigate a large obstacle). Anticipatory control is necessary to ensure that the robot avoids locomoting into a position/situation where it is unable to react.

6 Permissions

We have permission from Professor Howie Choset to use the "Snake Monster" Gazebo simulation environment for this project, and have proposed robot experiments if the simulated results appear promising.

References

[1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

[2] Philip Holmes, Robert J Full, Dan Koditschek, and John Guckenheimer. The dynamics of legged locomotion: Models, analyses, and challenges. *SIAM review*, 48(2):207–304, 2006.

[3] Auke Jan Ijspeert. Central pattern generators for locomotion control in animals and robots: a review. *Neural networks*, 21(4):642–653, 2008.

[4] Nate Kohl and Peter Stone. Policy gradient reinforcement learning for fast quadrupedal locomotion. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 3, pages 2619–2624. IEEE, 2004.

[5] Sanmit Narvekar, Jivko Sinapov, Matteo Leonetti, and Peter Stone. Source task creation for curriculum learning. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 566–574. International Foundation for Autonomous Agents and Multiagent Systems, 2016.

[6] Xue Bin Peng, Glen Berseth, and Michiel Van de Panne. Terrain-adaptive locomotion skills using deep reinforcement learning. *ACM Transactions on Graphics (TOG)*, 35(4):81, 2016.

[7] Ludovic Righetti and Auke Jan Ijspeert. Pattern generators with sensory feedback for the control of quadruped locomotion. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 819–824. IEEE, 2008.

[8] Serge Rossignol, Réjean Dubuc, and Jean-Pierre Gossard. Dynamic sensorimotor interactions in locomotion. *Physiological reviews*, 86(1):89–154, 2006.

[9] Guillaume Sartoretti, Samuel Shaw, Katie Lam, Naixin Fan, Matthew Travers, and Howie Choset. Central pattern generator with inertial feedback for stable locomotion and climbing in unstructured terrain. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–5. IEEE, 2018.

[10] Guillaume Sartoretti, Yunfei Shi, William Paivine, Matthew Travers, and Howie Choset. Distributed learning for the decentralized control of articulated mobile robots. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–6. IEEE, 2018.

[11] Matthew E Taylor and Peter Stone. Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10(Jul):1633–1685, 2009.

[12] Matthew E Taylor, Peter Stone, and Yaxin Liu. Transfer learning via inter-task mappings for temporal difference learning. *Journal of Machine Learning Research*, 8(Sep):2125–2167, 2007.

[13] Ryota Yamashina, Masafumi Kuroda, and Tetsuro Yabuta. Caterpillar robot locomotion based on q-learning using objective/subjective reward. In *System Integration (SII), 2011 IEEE/SICE International Symposium on*, pages 1311–1316. IEEE, 2011.