Reproduction of Meta Reinforcement Learning for Optimal Design of Legged Robots

Shen Yining：11911613 | Xu Zherui：12011230

Yong Zhitao：12110824 | Fan Wangzhuo：12111942

Qiao Kai：12211112 | Zhang Caomeng：12112223

26th March 2024

Proposed Project Title Summary

Our project mainly refers to the previous paper “Meta Reinforcement Learning for Optimal Design of Legged Robots”, attempting to reproduce the content of literature research. There are several problems centering robot design parameters: the correlation with the final output, the process of their determination, and the generalization of application. To solve these problems, we need to understand the method and algorithm for this paper and search some extra literatures about legged robots, machine learning, and Markov Decision Process (MDP).

The data and parameters are from the paper we referenced. A base design comes from a robot currently being developed, which consists of ANYmal C main body with longer legs. The nominal length is 350 mm for both thigh and shank links, on the basis of simplified considerations similar to those mentioned in. Other parameters have also been considered, such as the gear ratio of the actuators, the geometry of the linkage transmission, the attachment point of the legs to the base, and the orientation of the first actuator. In addition, a simplified model for the velocity and torque limits of the real actuator is included in the simulation. All the policies are trained for N=2000 epochs using the same hyper parameters for PPO. Each epoch runs 1000 training environments with random velocity commands and terrain parameters.

In terms of methods and algorithms, our project requires Markov Decision Process (MDP), Model-Agnostic Meta-Learning(MAML), and design optimization. MDP is a mathematical framework for formulating a discrete-time decision-making process which is commonly used in RL. The objective of RL is to obtain an optimal policy $\pi*$ that maximizes the cumulative discounted rewards throughout interactions with its environment in an iterative fashion. The MAML approach is used to train our meta-policies. This approach enables fast fine-tuning to a task with a small amount of data possible during test time. And design optimization is to obtain a set of design parameters that maximizes a given fitness function by using CMA-ES for the optimization.

We will evaluate our results by comparing them with the results and conclusions of previous papers.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What is the problem that you will be investigating?

"In the case of legged robots, these design parameters can include limb lengths, drive-train parameters such as gear ratio, and control parameters such as gait parameters and duration. The wide range of continuous and discrete design parameters results in a combinatorial problem with often unclear correlations between the design parameters and the resulting robot behavior…, but often it is unclear how the final values (of the parameters) are determined…since the motion parameters and simplified dynamics models are often developed/tuned for a certain instance by hand, it is hard to claim that the optimized motion is truly optimal for each design…"

There are several problems centering robot design parameters: the correlation with the final output, the process of their determination, and the generalization of application.

AncoraSIR.com

What reading will you examine?

To provide context and background

- Our project mainly refers to the previous paper “Meta Reinforcement Learning for Optimal Design of Legged Robots”, attempting to reproduce the content of literature research.

- In addition, we need to search literatures about legged robots, machine learning, and Markov Decision Process (MDP), to solve the problems we may encounter.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What data will you use?

1. A base design comes from a robot currently being developed, which consists of ANYmal C main body with longer legs. The nominal length is 350 mm for both thigh and shank links, on the basis of simplified considerations similar to those mentioned in.

2. Other parameters have also been considered, such as the gear ratio of the actuators, the geometry of the linkage transmission, the attachment point of the legs to the base, and the orientation of the first actuator.

3. In addition, a simplified model for the velocity and torque limits of the real actuator is included in the simulation.

4. All the policies are trained for N=2000 epochs using the same hyperparameters for PPO. Each epoch runs 1000 training environments with random velocity commands and terrain parameters.

AncoraSIR.com

What data will you use?

References:

[1] A. Ananthanarayanan et al., “Towards a bio-inspired leg design for high_x0002_speed running,” Bioinspiration & biomimetics, vol. 7, p. 046005, 08,2012.

[2] “Anymal - autonomous legged robot,” May 2021. [Online]. Available:https://www.anybotics.com/anymal-autonomous-legged-robot.

[3] J. Hwangbo et al., “Per-contact iteration method for solving contact dynamics,” IEEE Robotics and Automation Letters, vol. 3, no. 2, pp.895–902, 2018. [Online]. Available: www.raisim.com

Hyperparameter	Value
Discount factor γ	0.993
Entropy coefficient	0.0
Adam stepsize α	5×10^{-4}
GAE lambda λ	0.95
Clipping parameter	0.2
Meta-batch size M	5
Num. Mini-batches	10

PPO hyperparameters

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What method or algorithm are you proposing?

- **Markov Decision Process(MDP):** MDP is a mathematical framework for formulating a discrete-time decision-making process which is commonly used in RL. The objective of RL is to obtain an optimal policy $\pi*$ that maximizes the cumulative discounted rewards throughout interactions with its environment in an iterative fashion.

- **Fast Adaptation with Meta-learning:** The Model-Agnostic Meta-Learning (MAML) approach is used to train our meta-policies. This approach enables fast fine-tuning to a task with a small amount of data possible during test time.

- **Design Optimization:** The goal is obtaining a set of design parameters that maximizes a given fitness function by using CMA-ES for the optimization.

AncoraSIR.com

How will you evaluate your results?

We will evaluate our results by comparing them with the results and conclusions of previous papers.

TABLE I
OPTIMIZED LINK SCALES WITH RESPECT TO THE NOMINAL DESIGN

Objective	Flat		Easy Hills		Mid Hills		Hard Hills		Easy Steps		Mid Steps		Hard Steps	
	Thigh	Shank	Thigh	Shank	Thigh	Shank	Thigh	Shank	Thigh	Shank	Thigh	Shank	Thigh	Shank
$\mathcal{C}_v$	1.02	0.99	1.01	1.01	1.06	1.03	1.23	1.18	1.05	1.0	1.07	1.06	1.21	1.17
$\mathcal{C}_\tau$	0.61	0.63	0.63	0.67	0.64	0.68	0.75	0.80	0.70	0.68	0.76	0.77	0.94	0.97
$\mathcal{C}_p$	1.05	0.94	1.07	0.95	1.06	0.93	1.10	0.97	1.04	0.93	1.07	0.96	1.17	1.13

TABLE II
MEAN IMPROVEMENT IN OPTIMIZATION OBJECTIVES COMPARED TO THE NOMINAL DESIGN.

	$\mathcal{C}_v$	$\mathcal{C}_\tau$	$\mathcal{C}_p$
Flat	1.27%	43.53%	4.30%
Easy Hills	2.16%	43.85%	5.07%
Mid Hills	4.32%	39.72%	3.01%
Hard Hills	27.85%	16.36%	13.47%
Easy Steps	4.50%	37.47%	4.10%
Mid Steps	6.45%	28.98%	5.47%
Hard Steps	24.79%	4.13%	16.01%

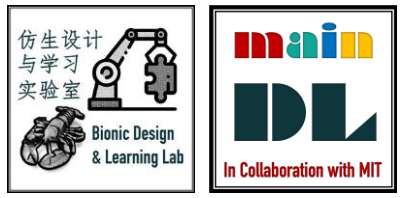
AncoraSIR.com

SUSTech
Southern University of Science and Technology

Thank you for your listening!

Team 2

ME 336 Spring 2024



AncoraSIR.com

SUSTech
Southern University
of Science and Technology