

Neuroevolution for Reinforcement Learning Using Evolution Strategies

Christian Igel

Institut für Neuroinformatik
Ruhr-Universität Bochum
44780 Bochum, Germany
christian.igel@neuroinformatik.rub.de

【强化学习 42】Neuroevolution



张楚珩

清华大学 交叉信息院博士在读

8 人赞同了该文章

随手找的一篇比较近的使用neuroevolution方法做强化学习的文章。

原文传送门

Salimans, Tim, et al. "Evolution strategies as a scalable alternative to reinforcement learning." arXiv preprint arXiv:1703.03864 (2017).

特色

大家来了解一下 Neuroevolution 吧！即结构和前向传播还是使用神经网络，但是参数优化就不用 BP 了，而是使用 evolution strategy (ES)。这篇文章就是使用这种方法来做 direct policy search。读完之后感觉做的工作并不是最前沿（考虑到这是2017年的工作），但是竟然有200+引用。

过程

1. 解决问题

文章主要针对 pole balancing 任务，主要分为了三个设定。

- 左右移动小车去平衡小车上面的一个杆，输入状态包括各种速度；
- 左右移动小车去平衡小车上面的两个杆，输入状态包括各种速度；
- 左右移动小车去平衡小车上面的两个杆，输入状态不包括各种速度，即状态不再是马可夫；

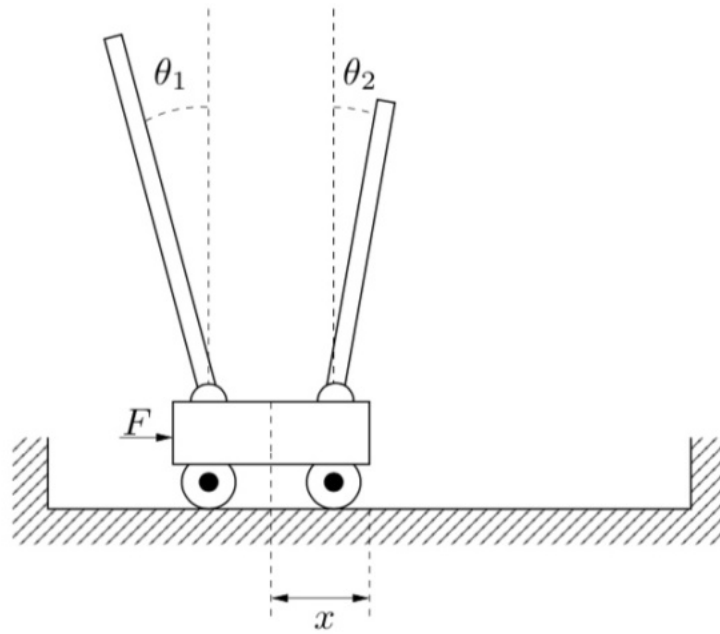


Figure 1: Double pole balancing problem. The parameters x , θ_1 , and θ_2 are the offset of the cart from the center of the track and the angles from the vertical of the long and short pole, respectively.

2. 方法

本问题的状态（state）为小车的位置、速度，杆的角度、角速度等；行动（action）为离散的把小车往左或者往右推；由于这个任务里面直接衡量最后的总奖励（fitness function），而不使用分步的奖励（reward）了，总奖励主要衡量保持平衡的时间和平衡后的晃动程度。

对于前两个设定，直接使用普通的单一隐含层MLP，对于后者使用了最为简化的RNN。

$$x_i(t) = f \left(\sum_{j=1}^{n_{\text{inputs}}} w_{ij} x_j(t) + \sum_{j=n_{\text{inputs}}+1}^{n_{\text{neurons}}} w_{ij} x_j(t-1) \right)$$

3. 实验结果

实验结果上能够比之前的基于ES的方法（其中某些方法更为复杂）使用更少的 attempts 找到一个能够使得小杆平衡的策略。

4. 其他细节和评述

- 由于平衡小车问题具有对称性，神经网络中去掉 bias 项，对于最后的结果有较明显的提升。
- 结构简单的神经网络更容易陷入局部极小值点。
- 针对CMA-ES中的mutation vector设定最低探索的方差，能够避免在选择压力（selection pressure）很小的时候停止探索，同时也能减轻陷入局部极小值点的问题。
- 本文优化的最大的神经网络参数也不超过200，仍然在CMA-ES舒适的维度范围内，没有证据表明ES方法能够应用于大规模神经网络，其拓展性是个大问题。

发布于 2019-03-03

强化学习 (Reinforcement Learning)

▲ 赞同 8 ▼

💬 2 条评论

🔗 分享

♥ 喜欢

★ 收藏

...

文章被以下专栏收录



强化学习前沿
读呀读paper

进入专栏