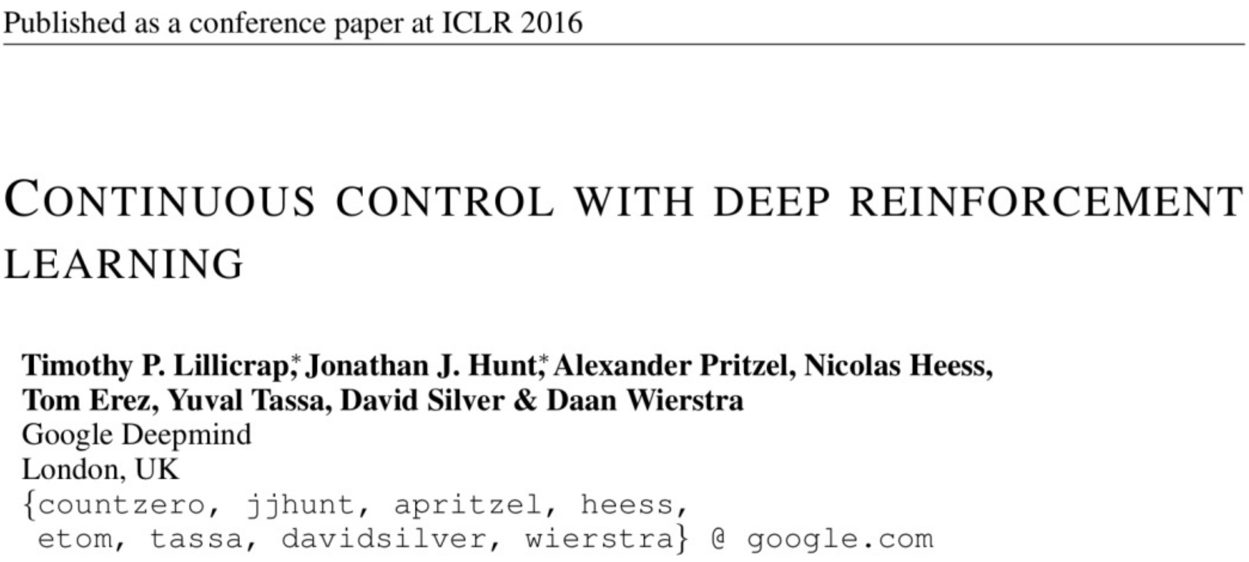

Published as a conference paper at ICLR 2016

Continuous control with deep reinforcement learning

Timothy P. Lillicrap[*], Jonathan J. Hunt[*], Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver & Daan Wierstra
Google Deepmind
London, UK
{countzero, jjhunt, apritzel, heess,
etom, tassa, davidsilver, wierstra} @ google.com

【强化学习算法 2】DDPG

张楚珩
清华大学 交叉信息院博士在读

5 人赞同了该文章

原文传送门：

Silver, David, et al. "Deterministic policy gradient algorithms." ICML. 2014.（前序工作）

Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015).

特色：能够处理连续行动空间的问题；使用了类似DQN的工程技巧使得本来很难稳定的off-policy+NN+bootstrap (actor-critic)问题能够运行。

分类：Model-free、Policy-based（actor-critic）、Off-policy、Continuous Action Space、Continuous State Space、Support High-dim Input

理论依据：Deterministic (off-policy) policy gradient theorem

$$\nabla_\theta J(\theta) \approx \mathbb{E}_\beta[\nabla_a Q_\varphi(s_t,a)|_{a=\mu(s_t)} \nabla_\theta \mu_\theta(s_t)]$$

更新公式：

$$\theta \leftarrow \theta + \alpha \nabla_a Q_\varphi(s_t,a)|_{a=\mu(s_t)} \nabla_\theta \mu_\theta(s_t)$$

$$\varphi \leftarrow \varphi + (r_t + \gamma Q_{\varphi'}(s_{t+1},\mu_{\theta'}(s_{t+1})) - Q_\varphi(s_t,a_t)) \nabla_\varphi Q_\varphi(s_t,a_t)$$

用到的其他技术：

1. action加上了Ornstein-Uhlenbeck process产生的噪声，用于更好的探索，因为本身是一个deterministic的策略，本身探索就不太够；
2. target network和double Q-network，用exponential moving average的方式更新策略和价值函数的网络作为target network；

编辑于 2018-09-19

强化学习 (Reinforcement Learning)

▲ 赞同 5 ▼ 添加评论 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