Continuous Control

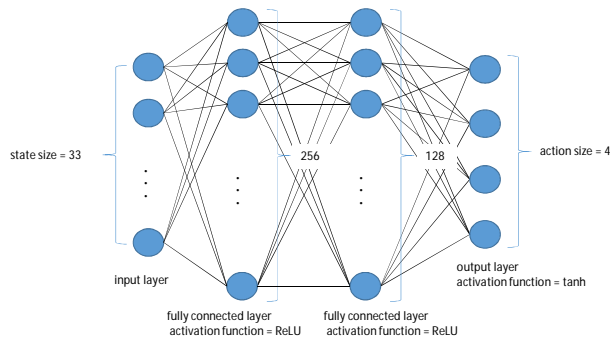Learning Argorithm   DDPG

Actor-Critic is a reinforcement learning method that learns by independently estimating the probability of an action and the estimated reward value of a state. DDPG (Deep Deterministic Policy Gradient) is an off-policy actor critic algorithm that combines DPG and DQN. DQN (Deep Q-Network) stabilizes the learning of Q-functions by using experience replay and fixing the target network. DQN works in discrete space, while DDPG is a continuous space algorithm.
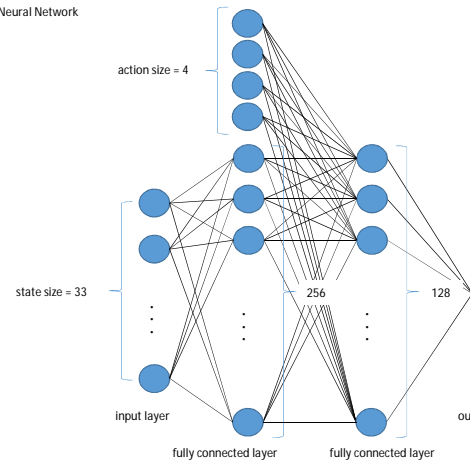
Hyperparameters

parameter			
		replay buffer size	100000
		batch size	128
		discount factor	0.99
		soft update of target parameters	0.001
	actor	learning rate	0.0001
	critic	learning rate	0.0001
		L2 weight decay	0.000001
		maximum number of training episodes	500
		maximum number of timesteps per episodes	1000
neural network		state size	33
		action size	4
	actor	number of nodes in first hidden layer	256
		number of nodes in second hidden layer	128
	ctitic	number of nodes in first hidden layer	256
		number of nodes in second hidden layer	128

Model Architecture

Actor Neural Network

state size = 33

256   128   action size = 4

input layer

output layer
activation function = tanh

fully connected layer   fully connected layer
activation function = ReLU   activation function = ReLU

Critic Neural Network

action size = 4

state size = 33

256   128

input layer

ou

fully connected layer   fully connected layer

Plot of Rewards

Reward values during training
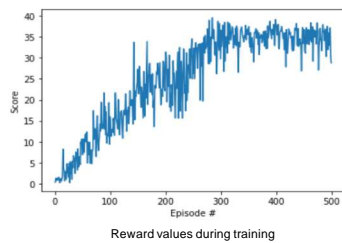
Ideas for Future Work

Like DQN, DDPG has been pointed out as having to overestimate action value. In TD3 (Twin Delayed DDPG) of the successor method, it is proposed to solve it by an approach similar to Double DQN.

utput layer