08

A2C 알고리즘

3. 결과분석

https://github.com/multicore-it/rl

결과분석

로그분석

```python
def build_model_actor(self):
    input_states = Input(shape=(self.state_size), name='input_states')
    input_action_matrixs = Input(shape=(self.action_size), name='input_action_matrixs')
    input_advantages = Input(shape=(self.value_size), name='input_advantages')

    x = (input_states)
    x = Dense(self.node_num, activation='relu')(x)
    out_actions = Dense(self.action_size, activation='softmax', name='output',
                        kernel_initializer='he_uniform')(x)

    model = self.MyModel(inputs=[input_states, input_action_matrixs,
                         input_advantages], outputs=out_actions)
    model.compile(optimizer=Adam(lr=self.learning_rate))

    model.summary()
    return model
```

Actor 모델 로그

Model: "my_model_3"

Layer (type)	Output Shape	Param #	Connected to
input_states (InputLayer)	[(None, 4)]	0	
dense_6 (Dense)	(None, 12)	60	input_states[0][0]
input_action_matrixs (InputLayer)	[(None, 2)]	0	
input_advantages (InputLayer)	[(None, 1)]	0	
output (Dense)	(None, 2)	26	dense_6[0][0]

Total params: 86
Trainable params: 86
Non-trainable params: 0

https://github.com/multicore-it/rl

결과분석

로그분석

```
def build_model_critic(self):
    input_states = Input(shape=(self.state_size), name='input_states')

    x = (input_states)
    x = Dense(self.node_num, activation='relu')(x)
    out_values = Dense(self.value_size, activation='linear', name='output')(x)

    model = tf.keras.models.Model(inputs=[input_states], outputs=[out_values])
    model.compile(optimizer=Adam(lr=self.learning_rate),
                  loss='mean_squared_error'
                 )
    model.summary()
    return model
```

Critic 모델 로그

Model: "my_model_7"

Layer (type)	Output Shape	Param #	Connected to
input_states (InputLayer)	[(None, 1, 4)]	0	
dense_7 (Dense)	(None, 1, 12)	60	input_states[0][0]
input_action_matrixs (InputLaye	[(None, 1, 2)]	0	
input_rewards (InputLayer)	[(None, 1, 1)]	0	
output (Dense)	(None, 1, 2)	26	dense_7[0][0]

Total params: 86
Trainable params: 86
Non-trainable params: 0

https://github.com/multicore-it/rl

결과분석

로그분석

A2C

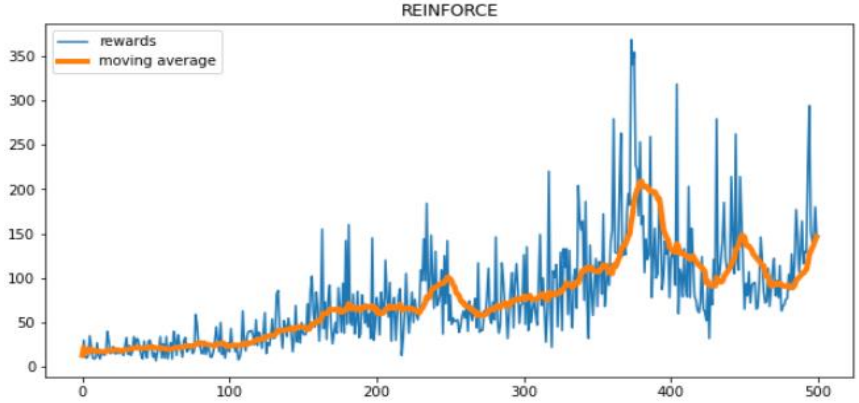
```
episode:270, moving_avg:51.3, rewards_avg:32.57933579335793
episode:280, moving_avg:58.0, rewards_avg:33.469750889679716
episode:290, moving_avg:88.0, rewards_avg:36.31958762886598
episode:300, moving_avg:129.7, rewards_avg:39.79734219269103
episode:310, moving_avg:205.85, rewards_avg:47.157556270096464
episode:320, moving_avg:278.5, rewards_avg:54.610591900311526
episode:330, moving_avg:222.55, rewards_avg:57.69788519637462
episode:340, moving_avg:197.05, rewards_avg:62.91202346041056
episode:350, moving_avg:243.1, rewards_avg:68.21367521367522
episode:360, moving_avg:222.85, rewards_avg:71.7202216066482
episode:370, moving_avg:176.7, rewards_avg:74.00808625336927
episode:380, moving_avg:207.6, rewards_avg:78.8005249343832
episode:390, moving_avg:233.1, rewards_avg:82.09462915601023
episode:400, moving_avg:164.15, rewards_avg:83.0074812967581
episode:410, moving_avg:75.3, rewards_avg:81.71532846715328
episode:420, moving_avg:27.2, rewards_avg:80.30878859857482
episode:430, moving_avg:29.15, rewards_avg:79.22969837587007
episode:440, moving_avg:164.45, rewards_avg:84.08843537414965
episode:450, moving_avg:397.05, rewards_avg:93.31042128603104
episode:460, moving_avg:500.0, rewards_avg:102.13232104121475
episode:470, moving_avg:381.8, rewards_avg:105.54352441613588
episode:480, moving_avg:192.9, rewards_avg:105.86902286902287
episode:490, moving_avg:131.4, rewards_avg:106.55600814663951
INFO:tensorflow:Assets written to: ./model/a2c\assets
*****end a2c learing
```

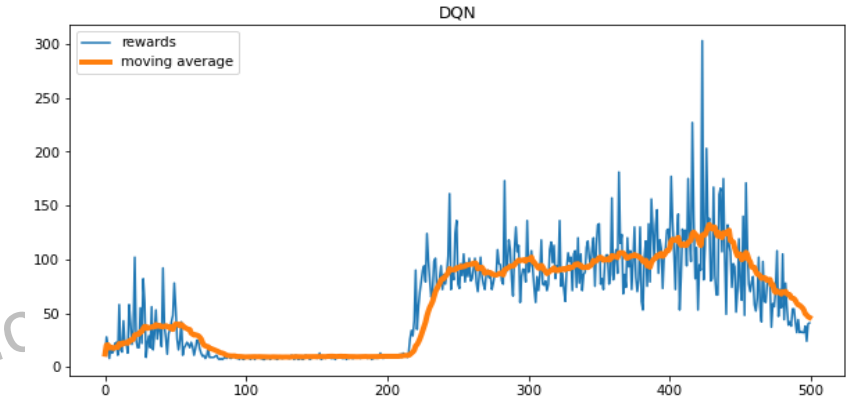
REINFORCE

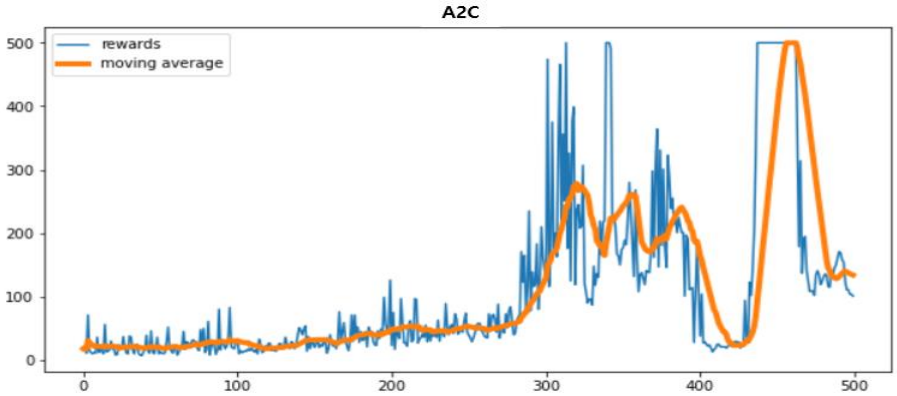
```
episode:270, moving_avg:88.75, rewards_avg:68.48708487084
episode:280, moving_avg:99.5, rewards_avg:69.44483985765
episode:290, moving_avg:156.4, rewards_avg:74.46048109965636
episode:300, moving_avg:179.75, rewards_avg:76.70764119601328
episode:310, moving_avg:128.95, rewards_avg:77.90032154340837
episode:320, moving_avg:112.25, rewards_avg:78.85981308411215
episode:330, moving_avg:112.85, rewards_avg:79.95166163141994
episode:340, moving_avg:120.75, rewards_avg:81.25806451612904
episode:350, moving_avg:171.9, rewards_avg:85.13390313390313
episode:360, moving_avg:199.8, rewards_avg:87.77008310249307
episode:370, moving_avg:180.25, rewards_avg:90.20754716981132
episode:380, moving_avg:161.25, rewards_avg:91.5748031496063
episode:390, moving_avg:127.15, rewards_avg:92.0460358056266
episode:400, moving_avg:115.75, rewards_avg:92.73067331670823
episode:410, moving_avg:131.75, rewards_avg:93.9294403892944
episode:420, moving_avg:177.2, rewards_avg:96.6959619952494
episode:430, moving_avg:188.3, rewards_avg:98.26218097447796
episode:440, moving_avg:129.9, rewards_avg:98.15646258503402
episode:450, moving_avg:119.9, rewards_avg:99.17738359201773
episode:460, moving_avg:161.8, rewards_avg:100.87418655097613
episode:470, moving_avg:149.95, rewards_avg:101.29087048832272
episode:480, moving_avg:139.4, rewards_avg:102.43451143451144
episode:490, moving_avg:207.3, rewards_avg:105.56822810590631
INFO:tensorflow:Assets written to: ./model/reinforce\assets
*****end learing
```

결과분석

시각화







https://github.com/multicore-it/rl