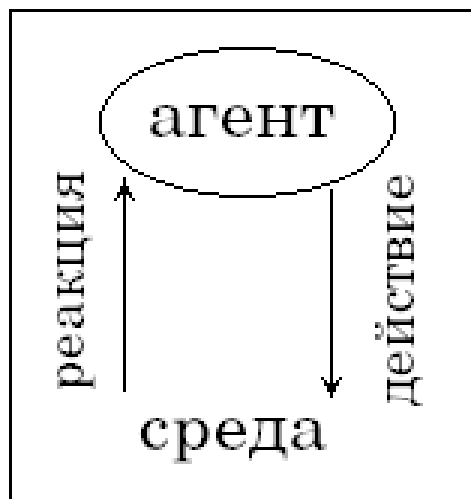




Обучение с подкреплением

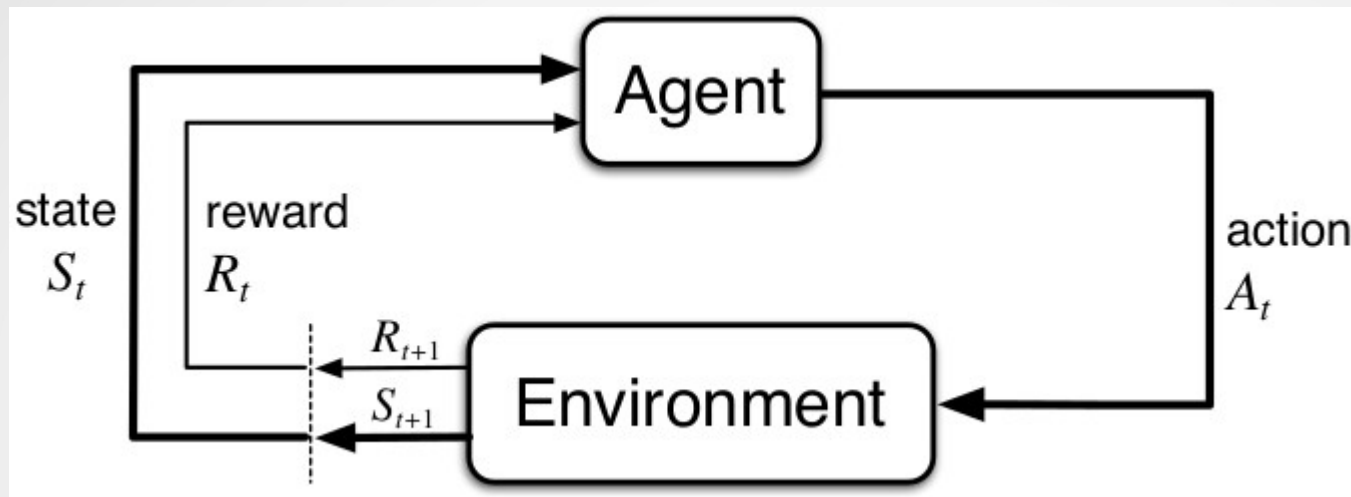
Евгений Борисов

ML: обучение с подкреплением



учебного набора в явном виде нет
собираем историю действий и последствий
пытаемся предсказывать реакцию среды
выбираем оптимальное действие

ML: обучение с подкреплением



S_t - состояние среды (state)

A_t - действие (action)

R_t - выигрыш (reward)

$\sum R_t$ суммарный выигрыш (cumulative reward)

$Q(S_t, A_t)$ - оценка суммарного выигрыша R в состоянии S_t от действия A_t

выбор оптимального действия $\Pi(S) = \operatorname{argmax}_a (Q(S, a))$

ML: обучение с подкреплением

Algorithm 1: deep Q-learning with experience replay.

Initialize replay memory D to capacity N

Initialize action-value function Q with random weights θ

Initialize target action-value function \hat{Q} with weights $\theta^- = \theta$

For episode = 1, M **do**

Initialize sequence $s_1 = \{x_1\}$ and preprocessed sequence $\phi_1 = \phi(s_1)$

For $t = 1, T$ **do**

With probability ε select a random action a_t

otherwise select $a_t = \operatorname{argmax}_a Q(\phi(s_t), a; \theta)$

Execute action a_t in emulator and observe reward r_t and image x_{t+1}

Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$

Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D

Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D

Set $y_j = \begin{cases} r_j & \text{if episode terminates at step } j+1 \\ r_j + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases}$

Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$ with respect to the network parameters θ

Every C steps reset $\hat{Q} = Q$

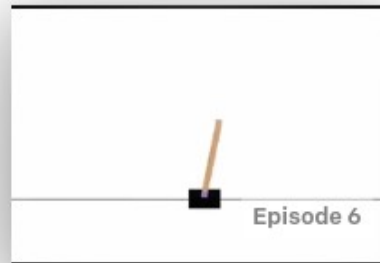
End For

End For

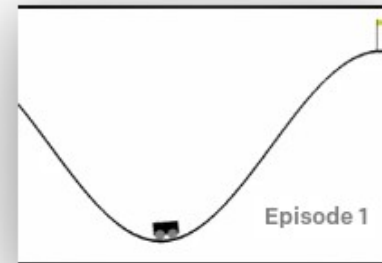
ML: обучение с подкреплением



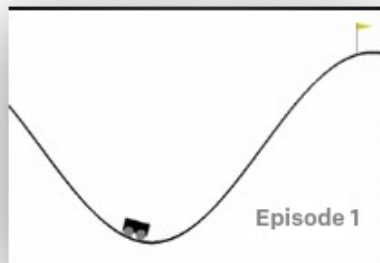
Acrobot-v1
Swing up a two-link robot.



CartPole-v1
Balance a pole on a cart.



MountainCar-v0
Drive up a big hill.

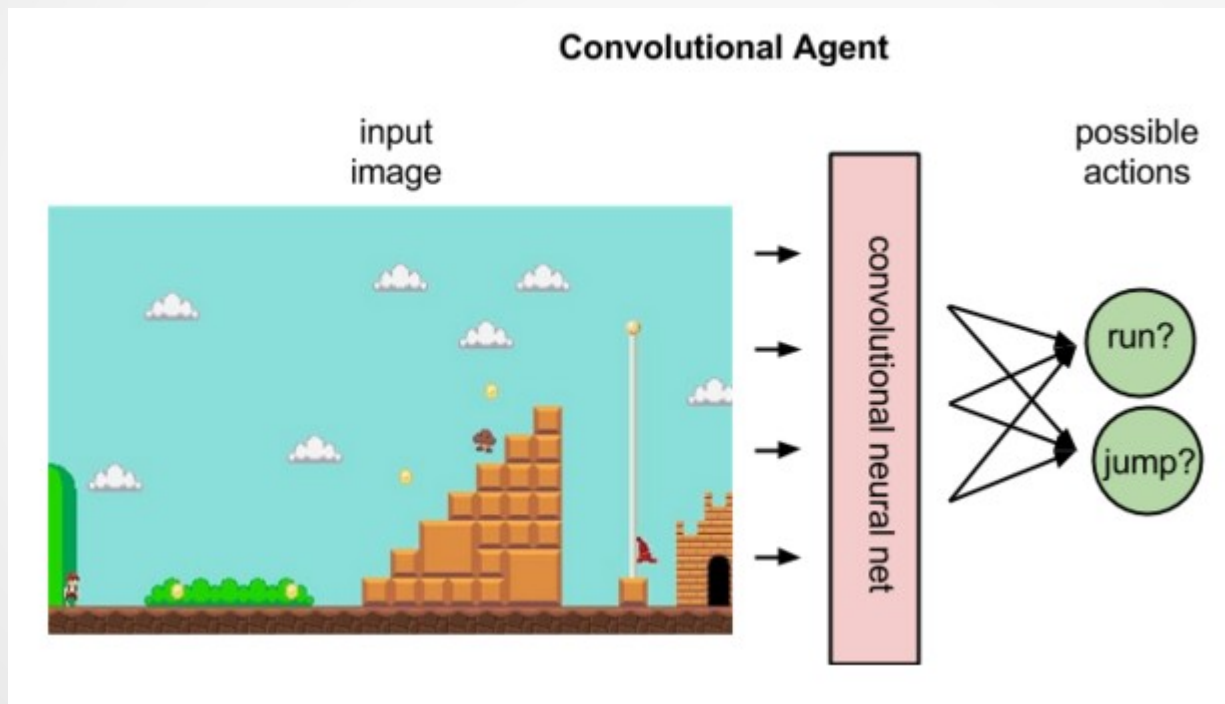


MountainCarContinuous-v0
Drive up a big hill with continuous control.



Pendulum-v0
Swing up a pendulum.

ML: обучение с подкреплением



<https://gym.openai.com>

обучение с подкреплением : литература

git clone https://github.com/mechanoid5/ml_lectorium.git

Николенко С., Кадури́н А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. - "Питер", 2018 г.

Саттон Р.С., Барто Э. Г. Обучение с подкреплением. - Москва:Бином, 2014г.