Reinforcement learning

FINAL PROJECT

Group member:
Wang Zeyao
Wang Chengzhi
Chen Chen

POLYTECH NICE-SOPHIA

CONTENTS

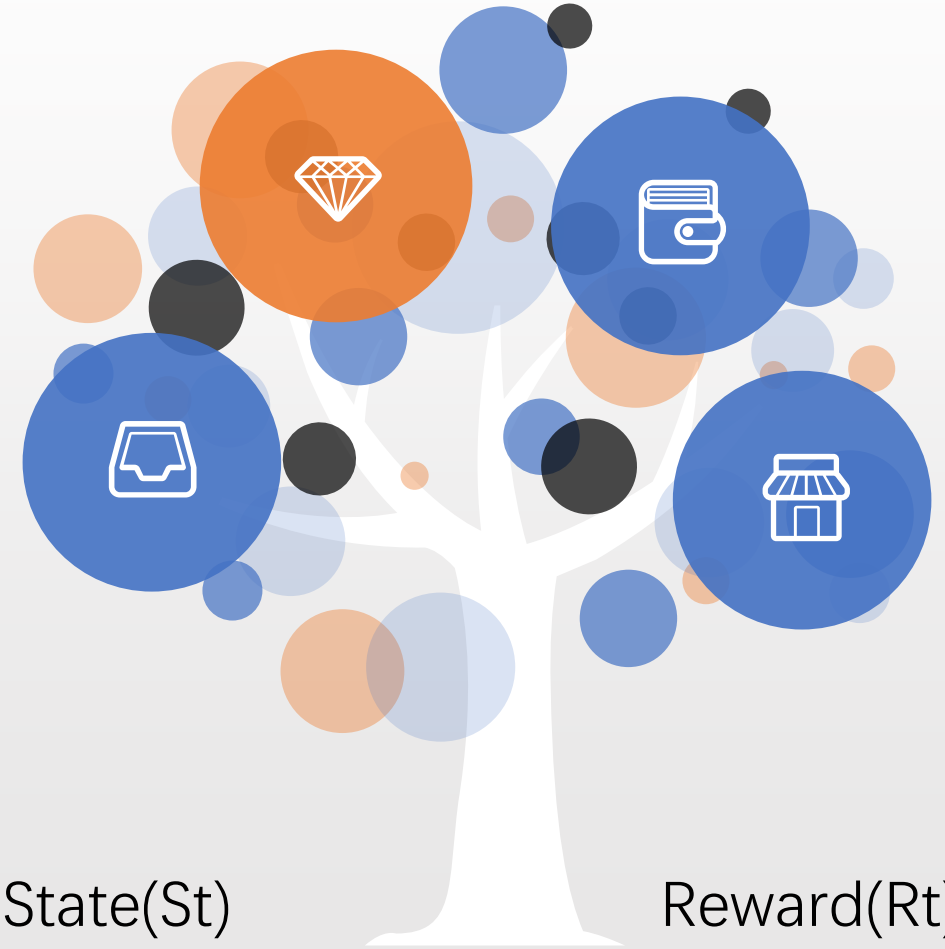
Introduction

Reinforcement learning

Agent

Action(At)

Discount factor



Q-value

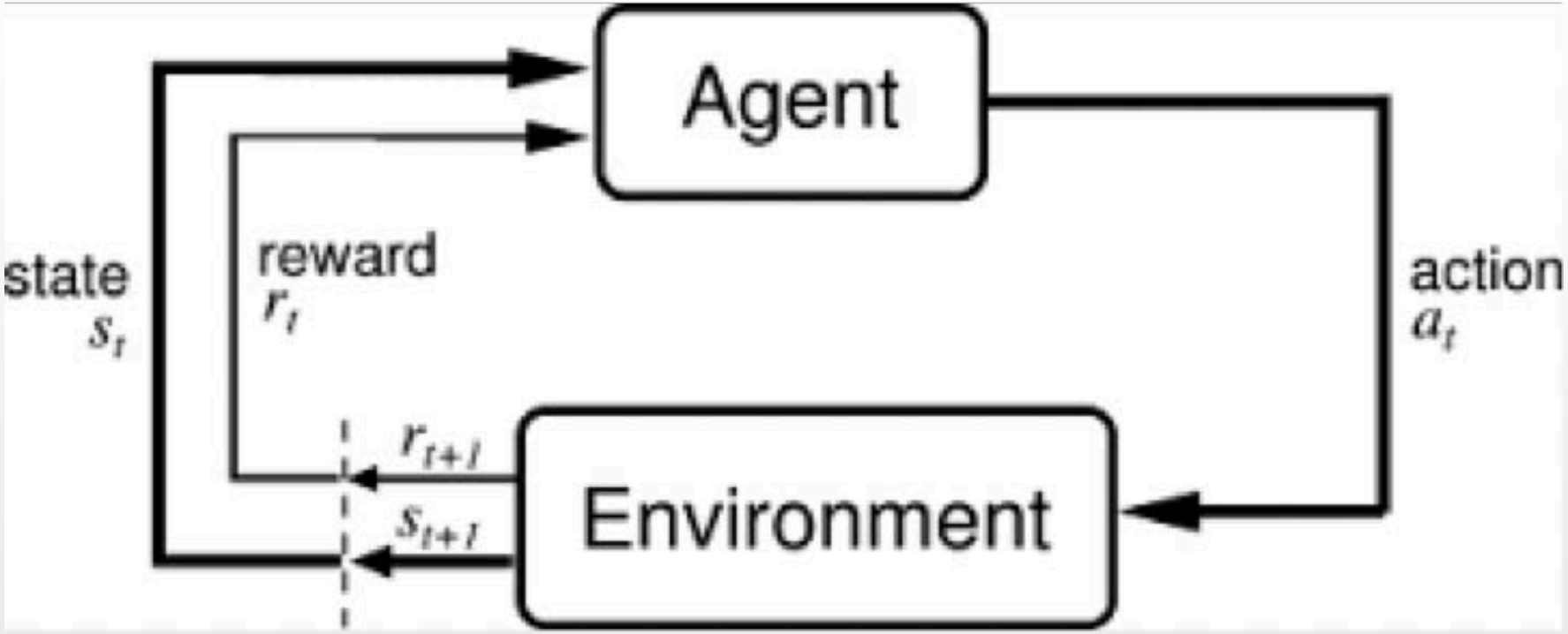
Value(Vt)

Policy

State(St)

Reward(Rt)

Reinforcement learning

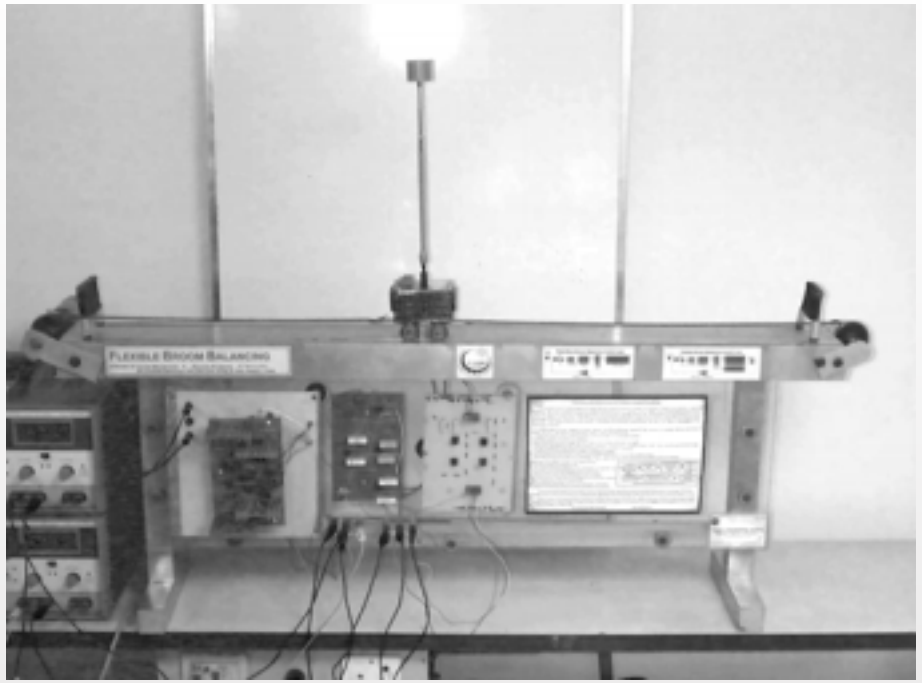


Markov chain

2

Physical method of cartpole problem

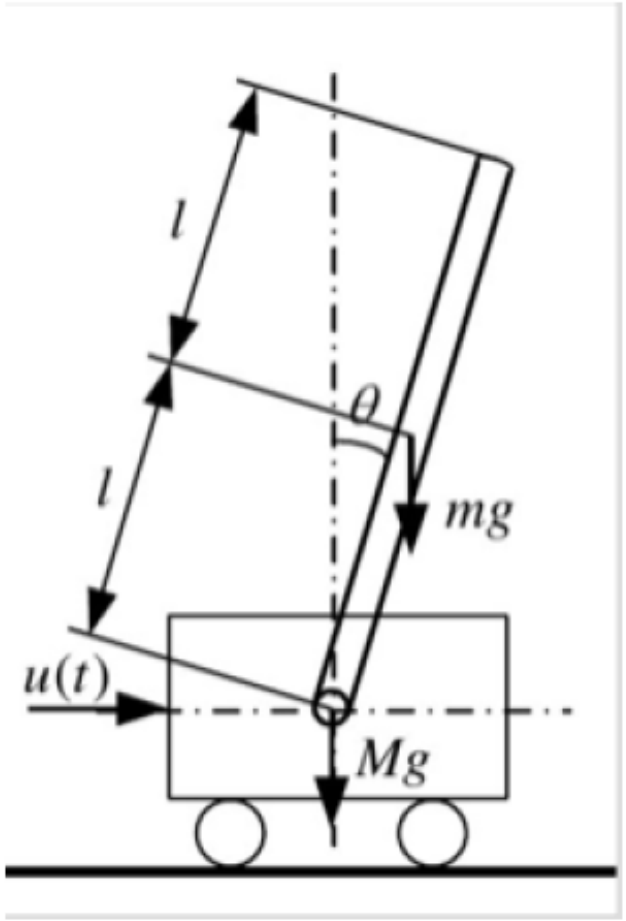
CARTPOLE PROBLEM IN LAB



A SIMPLE COUPLED SYSTEM

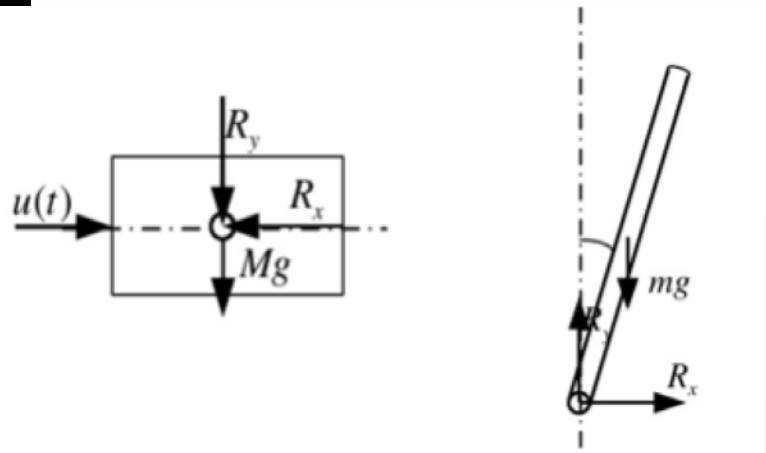
AN UNSTABLE SYSTEM

MATHEMATICAL MODEL



M	Mass of the Cart
m	Mass of the Pendulum
b	Friction of the Cart
L	Length of pendulum to Center of Gravity
I	Moment of Inertia (Pendulum)
R	Radius of Pulley,
τ_M	Time Constant of motor
K_M	Gain of Motor
K_F	Gain of Feedback
F	Force applied to the cart
x	Cart Position Coordinate
θ	Pendulum Angle with the vertical

MATHEMATICAL MODEL



$$\frac{\theta(s)}{U(s)} = \frac{3}{s^2 - 29.4}$$

$$M\ddot{x} = u(t) - R_x \quad (1)$$

$$m\ddot{x} + ml\ddot{\theta}cos\theta = R_x \quad (2)$$

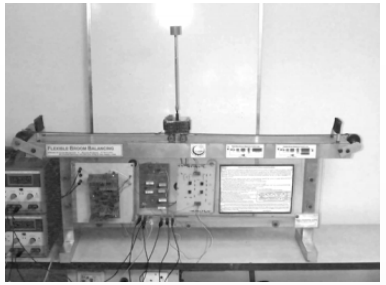
$$J\ddot{\theta} = mglsin\theta - ml\ddot{\theta}l - m\ddot{x}lcos\theta \quad (3)$$

$$sin\theta = \theta, cos\theta = 1, \theta^2 = 0$$

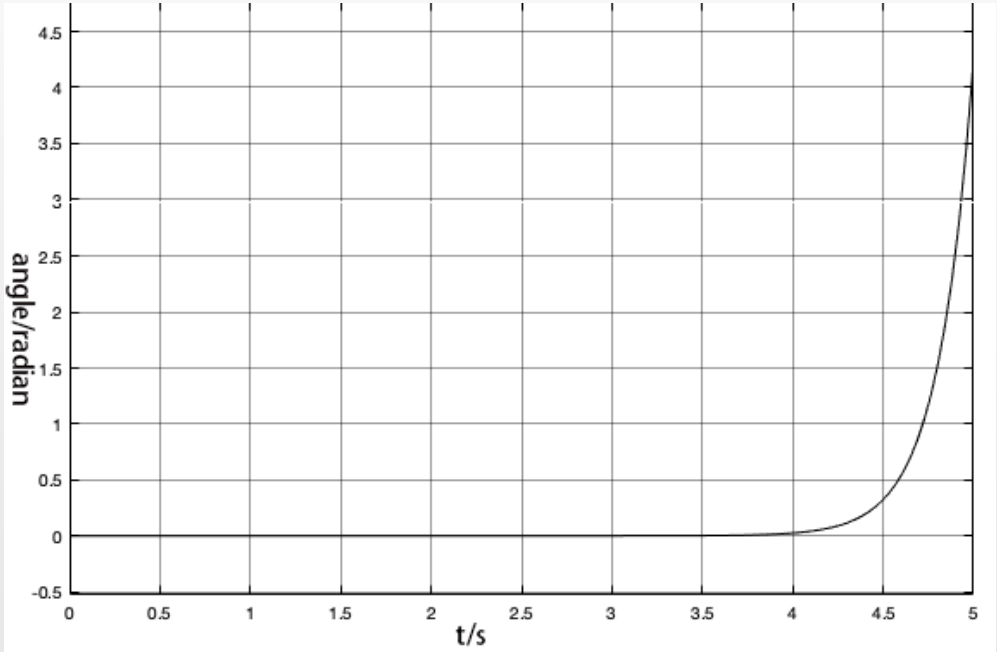
$$(M + m)\ddot{x} + ml\ddot{\theta} = u(t) \quad (4)$$

$$(J + ml^2)\ddot{\theta} + ml\ddot{x} = mgl\theta \quad (5)$$

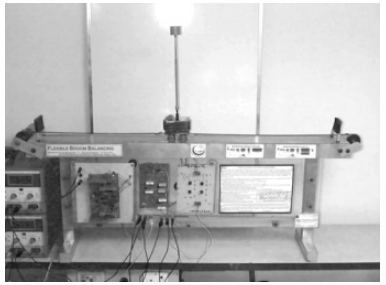
PID CONTROLLER



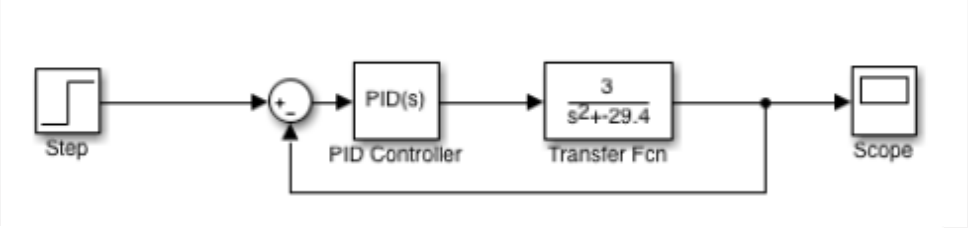
$$\frac{\theta(s)}{U(s)} = \frac{3}{s^2 - 29.4}$$

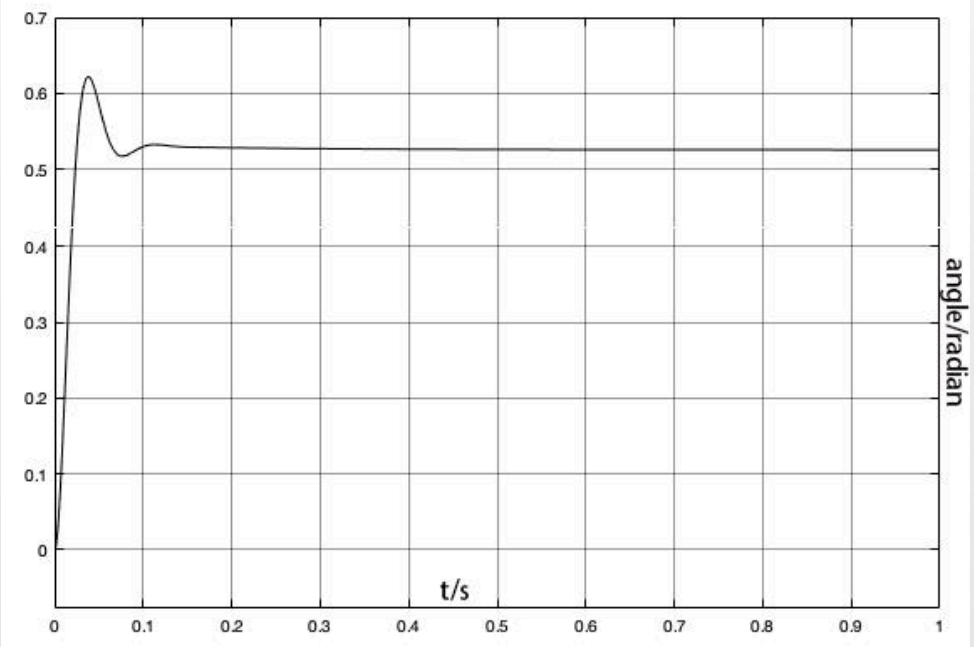


PID CONTROLLER



P:proportion
I: integral
D:differential





$$G(s) = \frac{R(s)}{E(s)} = K_P + \frac{K_I}{s} + K_D s = \frac{K_D s^2 + K_P s + K_I}{s}$$

KD=30
KP =600
KI=5000
t=0.18s

3 Introduction of Q-learning

Environment

POLYTECH NICE-SOPHIA

Environment:OpenAI gym cartpole V0

Observation

Type: Box(4)

Num	Observation	Min	Max
0	Cart Position	-2.4	2.4
1	Cart Velocity	-Inf	Inf
2	Pole Angle	~ -41.8°	~ 41.8°
3	Pole Velocity At Tip	-Inf	Inf

Environment

POLYTECH NICE-SOPHIA

Actions

Type: Discrete(2)

Num	Action
0	Push cart to the left
1	Push cart to the right

Reward

Reward is 1 for every step taken, including the termination step

Starting State

All observations are assigned a uniform random value between ±0.05

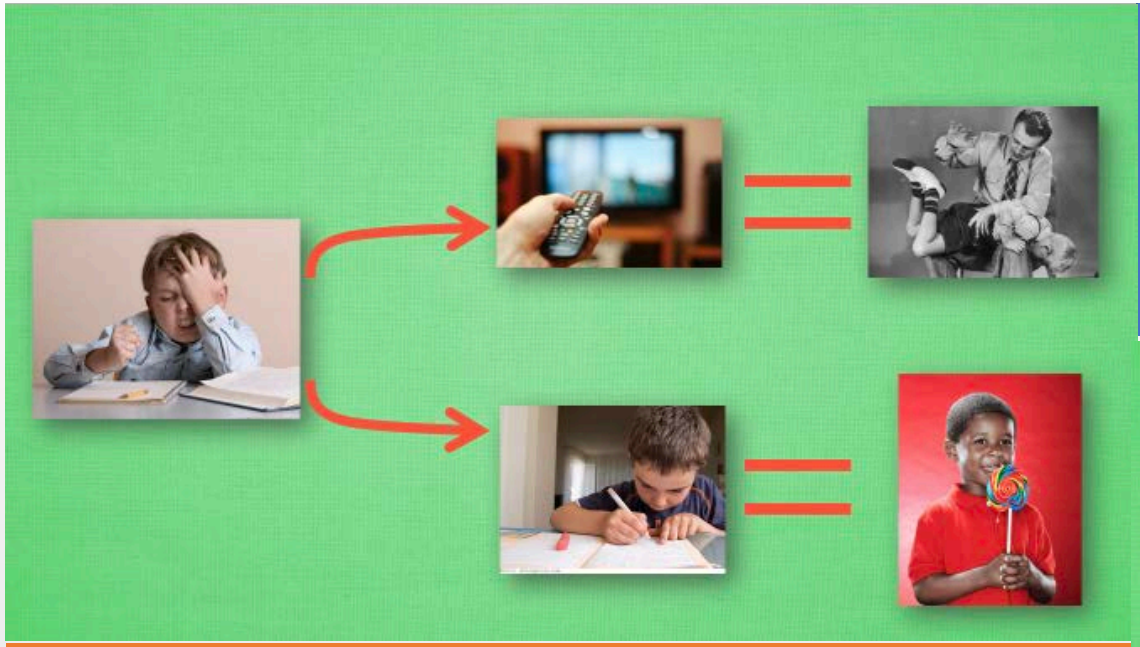
Episode Termination

1. Pole Angle is more than ±12°
2. Cart Position is more than ±2.4 (center of the cart reaches the edge of the display)

Solved Requirements

Considered solved when the average reward is greater than or equal to 195.0 over 100 consecutive trials.

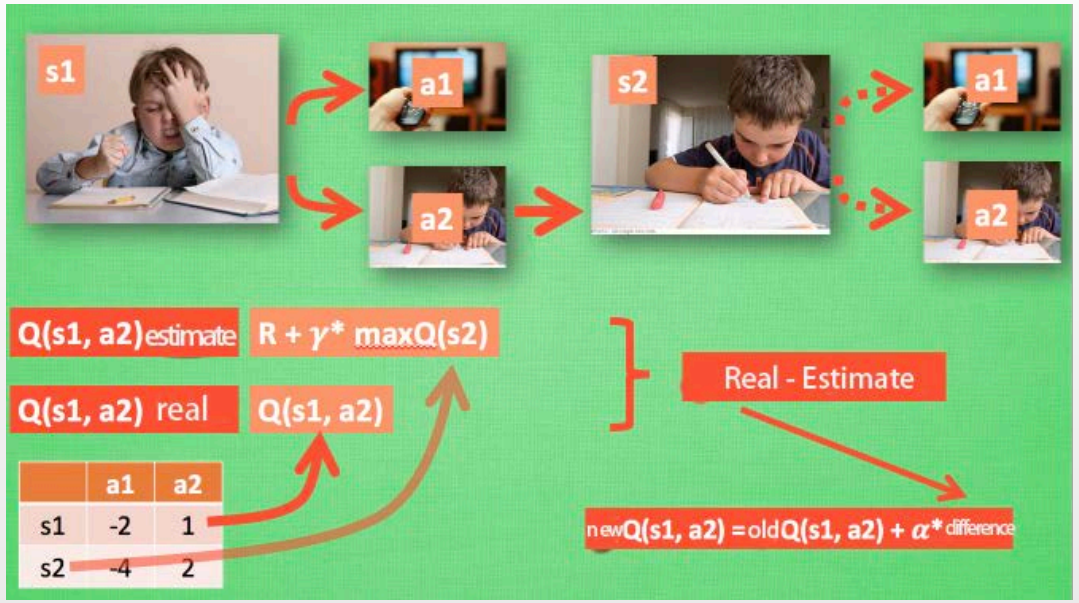
Process of Q-learning



Two choices:
(1) Continue to work -> get the praise
(2) Go to watch tv -> be punished

Experience -> potential reward
Choose the better one

Process of Q-learning

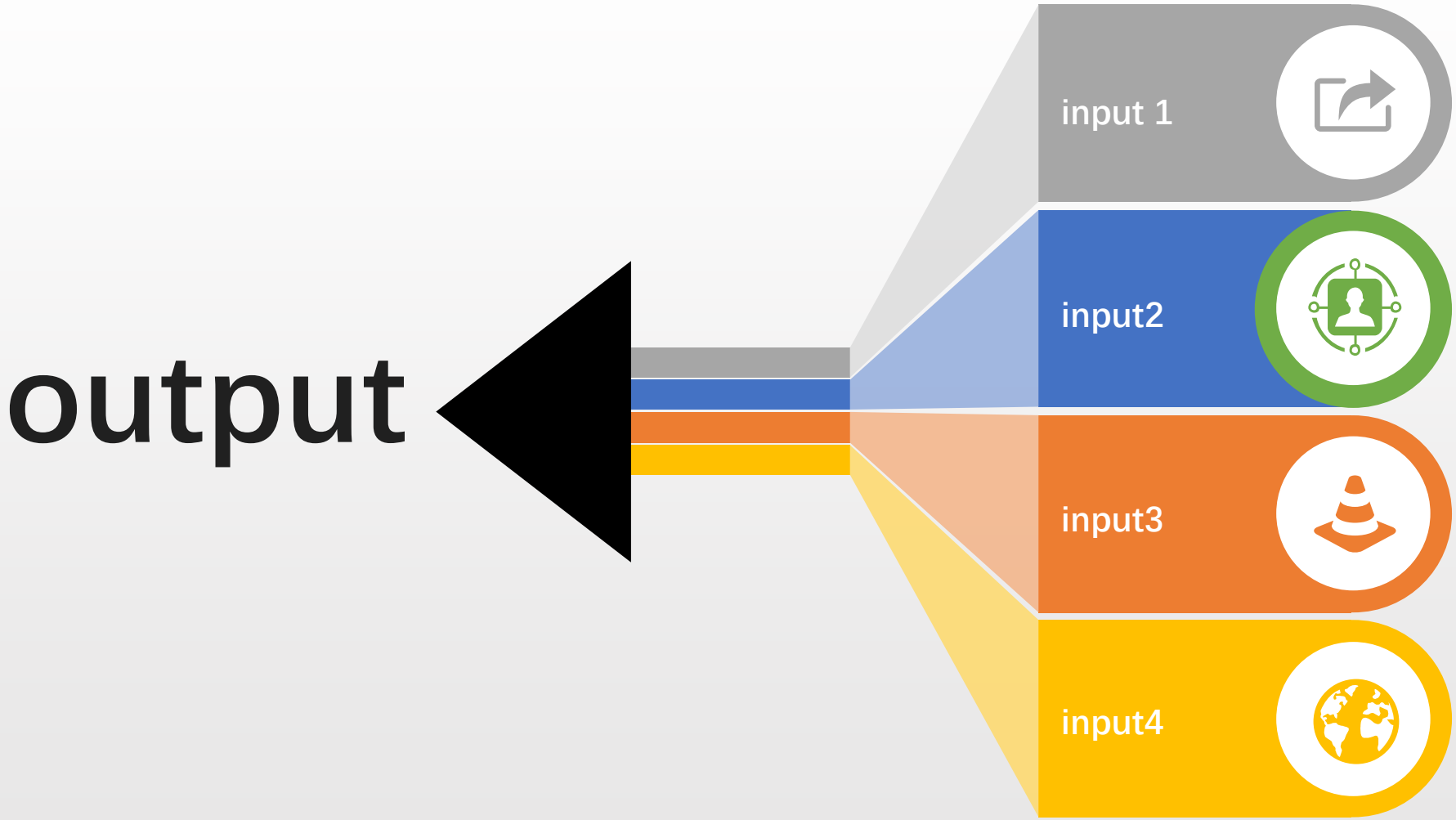


Experience => potential reward

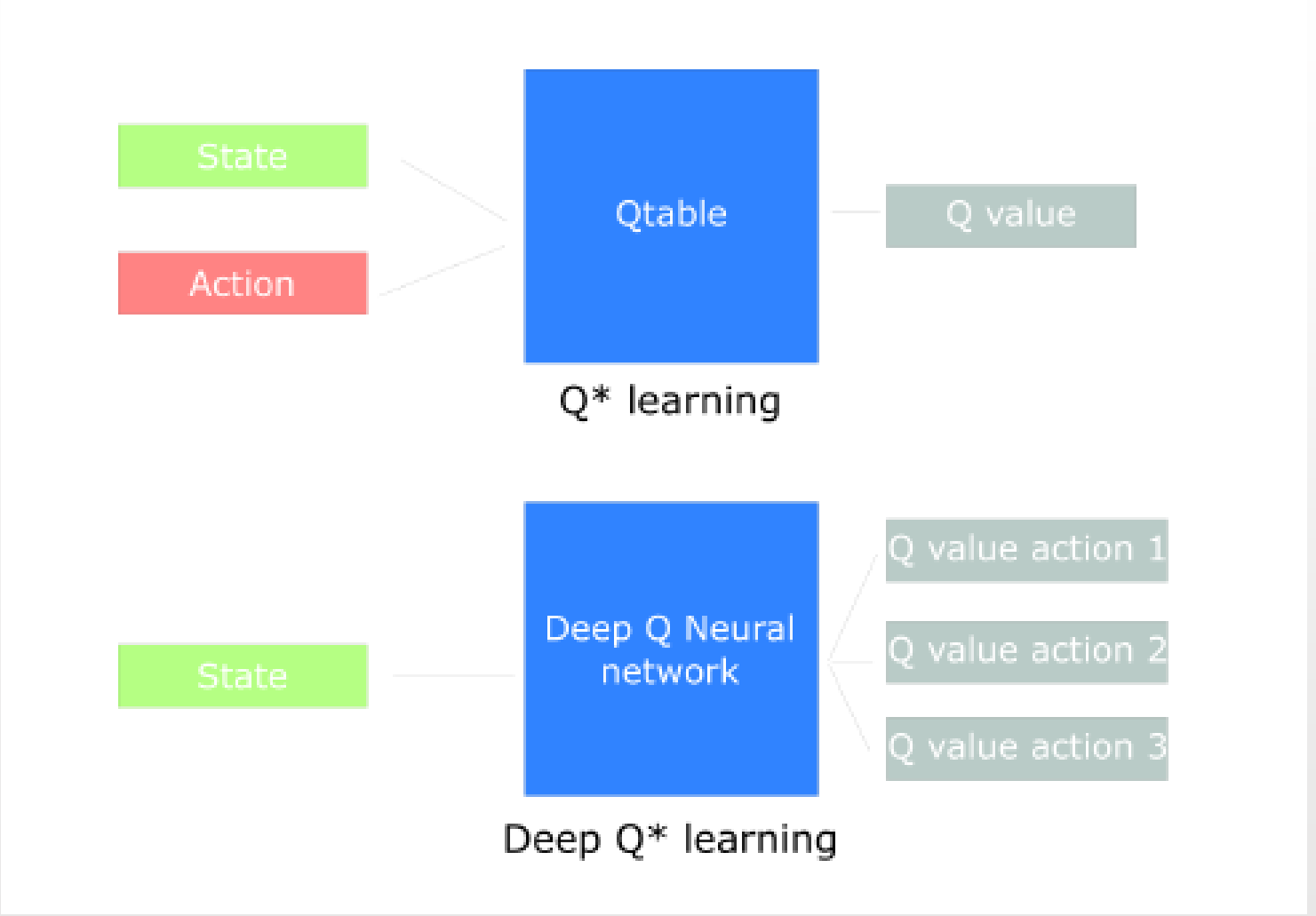
Initialize $Q(s, a)$ arbitrarily
Repeat (for each episode):
  Initialize s
  Repeat (for each step of episode):
    Choose a from s using policy derived from Q (e.g., ε-greedy)
    Take action a, observe r, s'
    $Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]$
    $s \leftarrow s'$;
  until s is terminal

Deep Q-learning

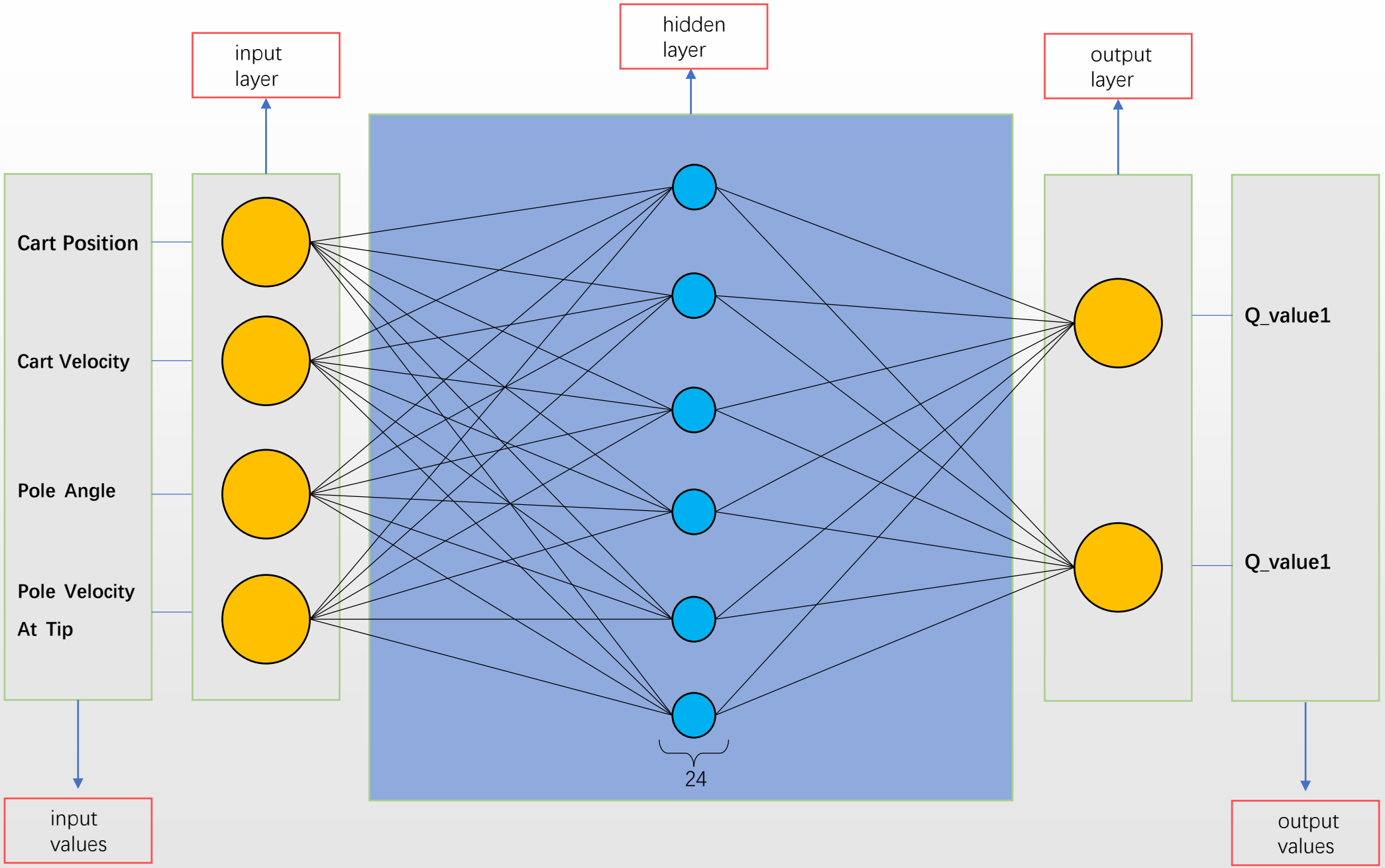
Neural network



Deep Q-learning



Deep Q-learning



5 Comparison between two methods

Condition and setup

POLYTECH NICE-SOPHIA

01

Physical condition

Gravity = 9,8 m/s2
Mass cart = 1.0kg
Mass pole = 0.1kg
Length pole = 0.25m

02

Initial conditions

CartpoleV0: random float between -0.05~0.05 for each variable in the observation space.
PID control: an instance acceleration of 0.05m/s2.

03

Control

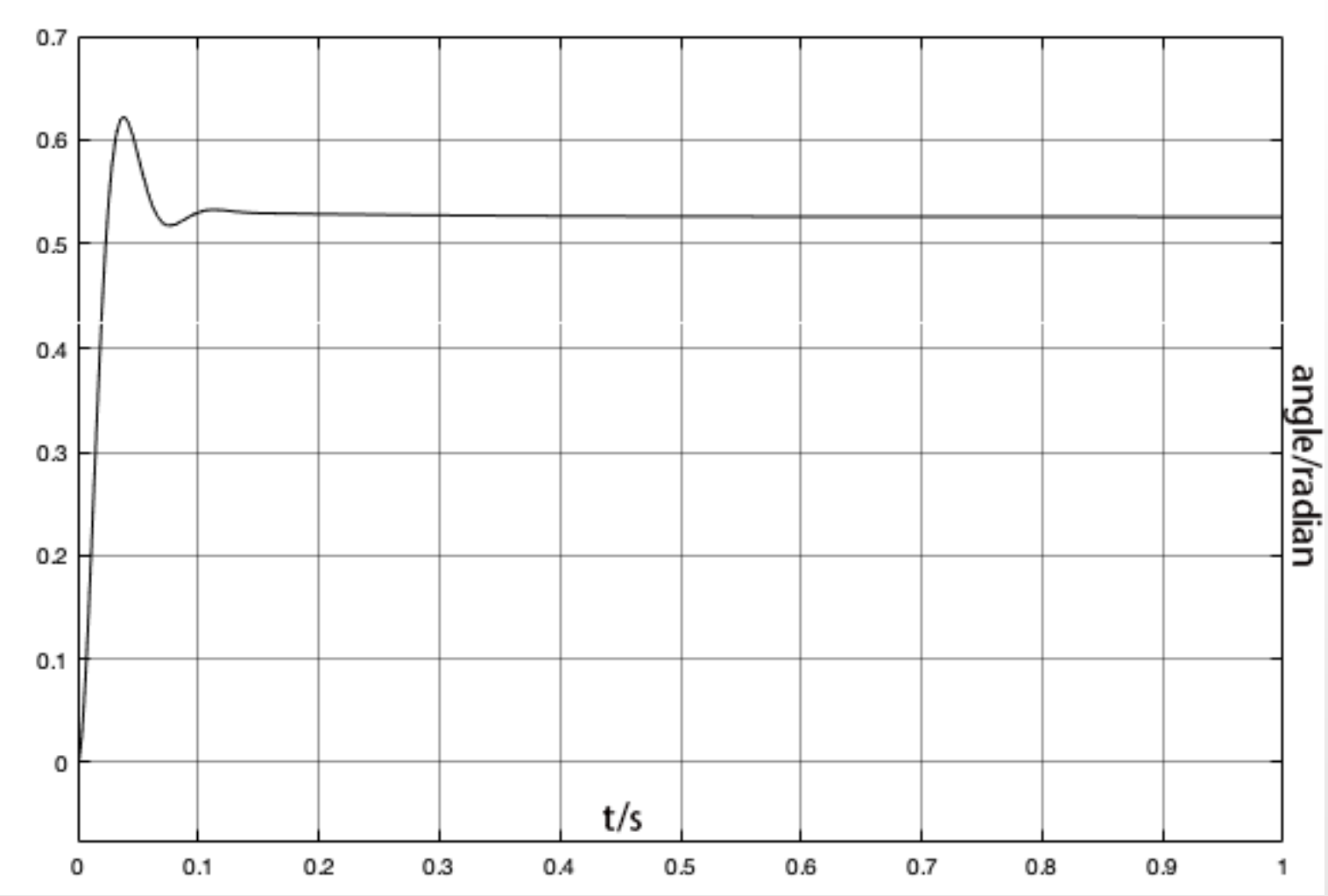
CartpoleV0: 10N OR -10N
PID control: $\frac{\theta(s)}{U(s)} = \frac{3}{s^2 - 29.4}$

04

Failure condition

CartpoleV0: Pole Angle is more than ± 0.1 radian
PID control: Pole radian can't converge.

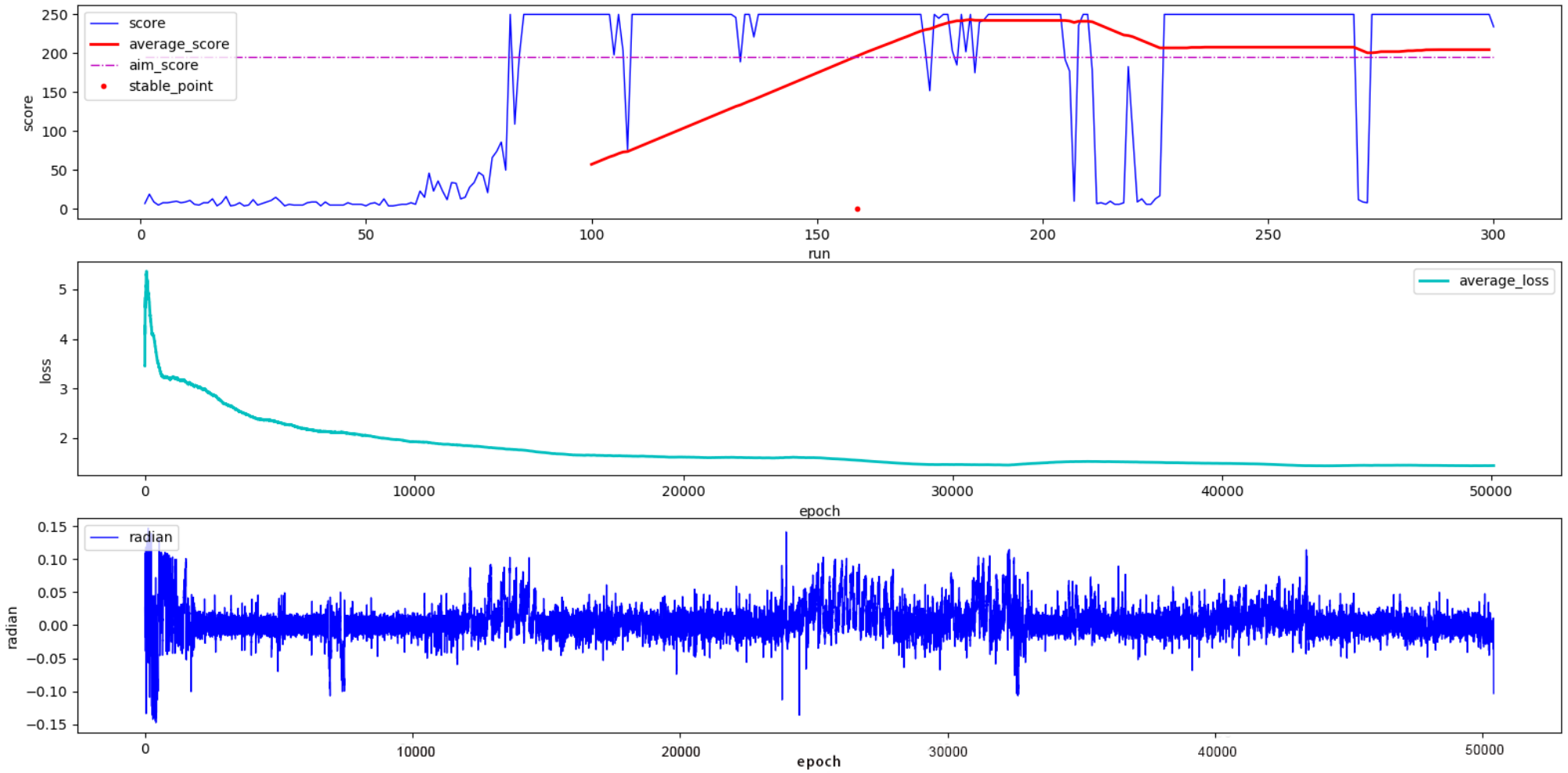
PID Control



Deep Q-learning

POLYTECH NICE-SOPHIA

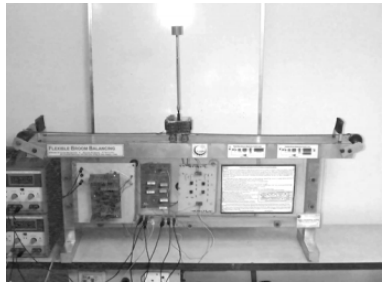
```
Stable time:  402.24s
Stable run: 159
```



Conclusion

- Advantages of Physical Method:
  - ✓ Stability
  - ✓ Less time spent to achieve stability
- Disadvantages of Physical Method:
  - ✓ Lots of preparations for the experiment
  - ✓ Large amounts of manual calculation
  - ✓ Parameters needed to be adjusted by practical situation and human experiment.
- Advantages of Reinforcement Learning Method:
  - ✓ Little preparation for experiment
  - ✓ Little calculation.
- Disadvantages of Reinforcement Learning Method:
  - ✓ Lack of Stability

Conclusion



Tanks