



VietAI

Introduction to Deep Re-inforcement Learning



Presenter: Bảo Đại



Nội dung

1. Giới thiệu về re-inforcement learning
2. Động lực sử dụng deep re-inforcement learning
3. Bài toán sử dụng deep re-inforcement learning
4. Sử dụng deep re-inforcement learning như thế nào
5. Một số công trình về deep re-inforcement learning



Nội dung

- 1. Giới thiệu về re-inforcement learning**
2. Động lực sử dụng deep re-inforcement learning
3. Bài toán sử dụng deep re-inforcement learning
4. Sử dụng deep re-inforcement learning như thế nào
5. Một số công trình về deep re-inforcement learning

The screenshot shows the Google Translate web interface. At the top, there are language selection buttons for 'Tiếng Anh', 'Tiếng Việt', 'Tiếng Pháp', and a dropdown for 'Phát hiện ngôn ngữ'. On the right, there are buttons for 'Tiếng Việt', 'Tiếng Anh', 'Tiếng Trung (Giản Thể)', and a blue 'Dịch' button. The input text 'reinforce' is entered in the left box, and the output 'củng cố' is shown in the right box. Below the input box, there is a section titled 'Nghĩa của reinforce' with a definition and a list of synonyms. To the right of this, there is a section titled 'Bản dịch của reinforce' with a list of synonyms. At the bottom of the page, there is a large, light gray chevron pointing downwards.

reinforce

9/5000

.rẽ-in' fòrs

Nghĩa của reinforce

đồng từ

strengthen or support, especially with additional personnel or material.
"paratroopers were sent to reinforce the troops already in the area"
từ đồng nghĩa: augment, increase, add to, supplement, boost, top up

Xem thêm
reinforce

Bản dịch của reinforce

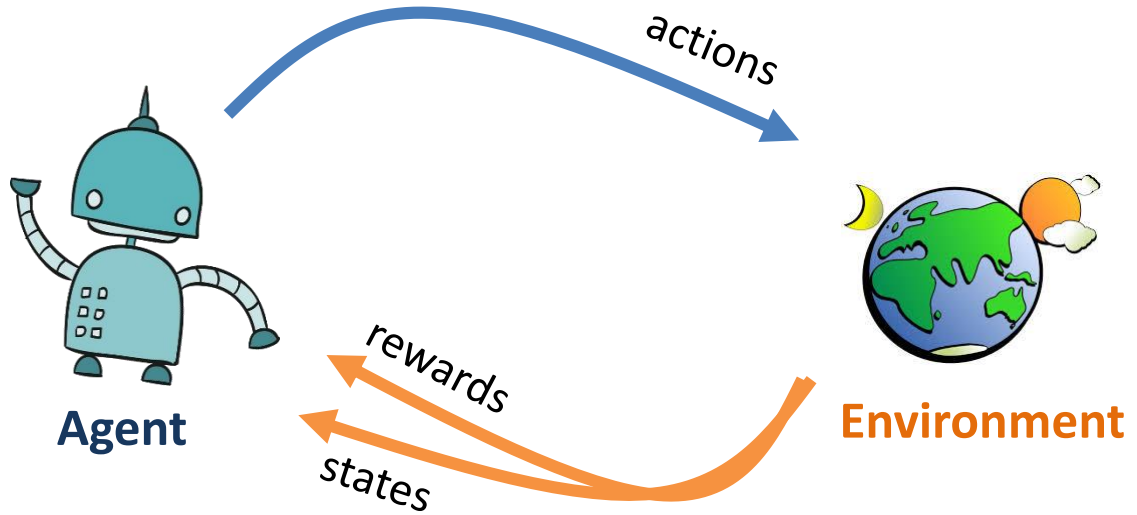
đồng từ

kiến cố thêm	reinforce
làm cho đồng hơn	reinforce
làm cho mạnh thêm	nerve, reinforce
vững thêm	reinforce

- Là một nhánh thuộc Machine learning
- Giúp máy (**agent**) có thể tự động đưa ra quyết định tốt nhất (**ideal behavior**) trong một ngữ cảnh nhất định
- Cần có “phần thưởng” để agent điều chỉnh hành vi

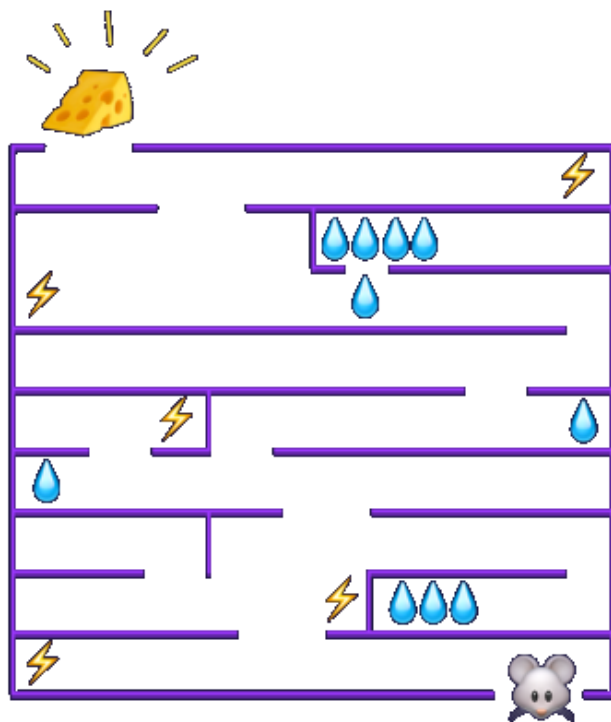


Giới thiệu RE





Giới thiệu RE



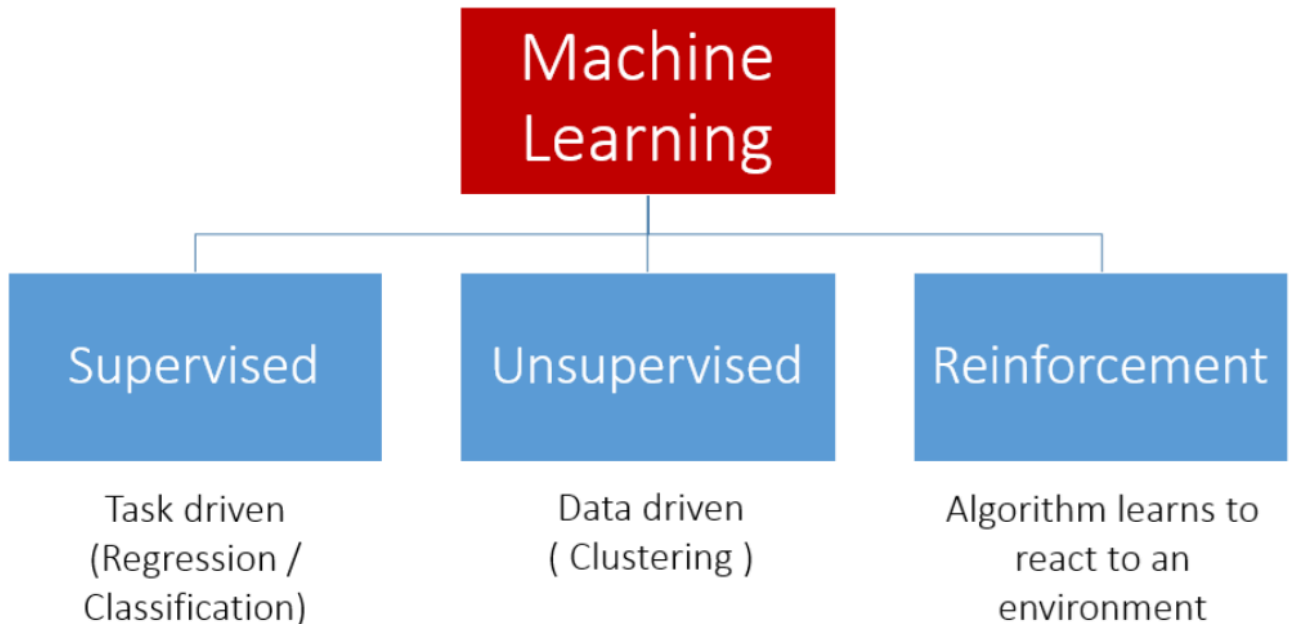


Giới thiệu RE





Types of Machine Learning





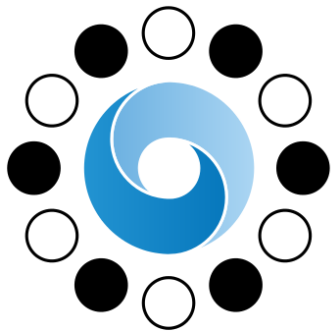
Giới thiệu Deep RE



DeepMind



Google DeepMind



AlphaGo



Nội dung

1. Giới thiệu về re-inforcement learning
- 2. Động lực sử dụng deep re-inforcement learning**
3. Bài toán sử dụng deep re-inforcement learning
4. Sử dụng deep re-inforcement learning như thế nào
5. Một số công trình về deep re-inforcement learning

- “Nhãn” được sinh ra dựa trên environment thay vì được gán công bởi chuyên gia trong domain
- Tiết kiệm thời gian chuẩn bị dữ liệu

...ions and selected moves using deep neural networks. These neural networks were trained by supervised learning from human expert moves, and by reinforcement learning from self-play. Here, we introduce an algorithm based solely on reinforcement learning, without human data, guidance, or domain knowledge beyond game rules. *AlphaGo* becomes its own teacher: a neural network is trained to predict *AlphaGo*'s own move selections and also the winner of *AlphaGo*'s games. This neural network improves the strength of tree search, resulting in higher quality move selection and stronger self-play in the next iteration. Starting



Nội dung

1. Giới thiệu về re-inforcement learning
2. Động lực sử dụng deep re-inforcement learning
- 3. Bài toán sử dụng deep re-inforcement learning**
4. Sử dụng deep re-inforcement learning như thế nào
5. Một số công trình về deep re-inforcement learning



- Playing Atari games

Volodymyr Mnih Koray Kavukcuoglu David Silver Alex Graves Ioannis Antonoglou Daan Wierstra Martin Riedmiller “**Playing Atari with Deep Reinforcement Learning**” – DeepMind Technology (2013)

- Computer vision

- Image captioning
- Object detection
- Action detection

- Natural language processing

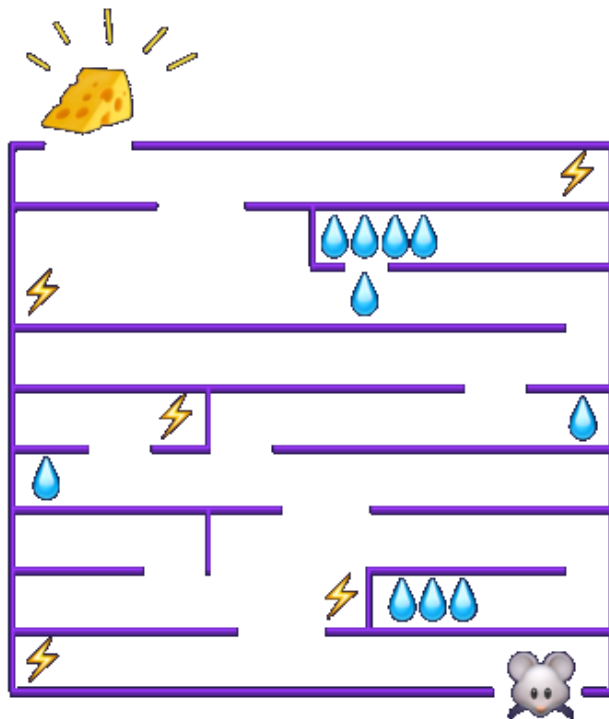
- Dialog Generation
- Information Extraction




Nội dung

1. Giới thiệu về re-inforcement learning
2. Động lực sử dụng deep re-inforcement learning
3. Bài toán sử dụng deep re-inforcement learning
- 4. Sử dụng deep re-inforcement learning như thế nào**
5. Một số công trình về deep re-inforcement learning

1. **Tập trạng thái S (set of states):** Những vị trí có thể có của agent
2. **Tập hành động A (set of actions):** Những hành động có thể đưa ra trong mỗi trạng thái nhất định
3. **Tập khả năng θ (transitions between states):** Xác suất trạng thái mới có thể đạt được sau khi thực hiện hành động
4. **Tập phần thưởng R (rewards):** ứng với mỗi transition
5. **Hệ số giảm γ ($0 \leq \gamma \leq 1$):** dùng để giảm reward tương lai
6. Tính chất “quên”: tương lai không phụ thuộc vào quá khứ, chỉ phụ thuộc vào state hiện tại



Q-table (Bảng trạng thái-hành động)



	a^1	a^2	a^3	...	a^m
s^1	0.23	0.81	0.4	...	0.54
s^2	0.1	0.54	0.97	...	0.2
s^3	0.3	0.81	0.76	...	0.3
...
s^n	0.4	0.4	0.76	...	0.76

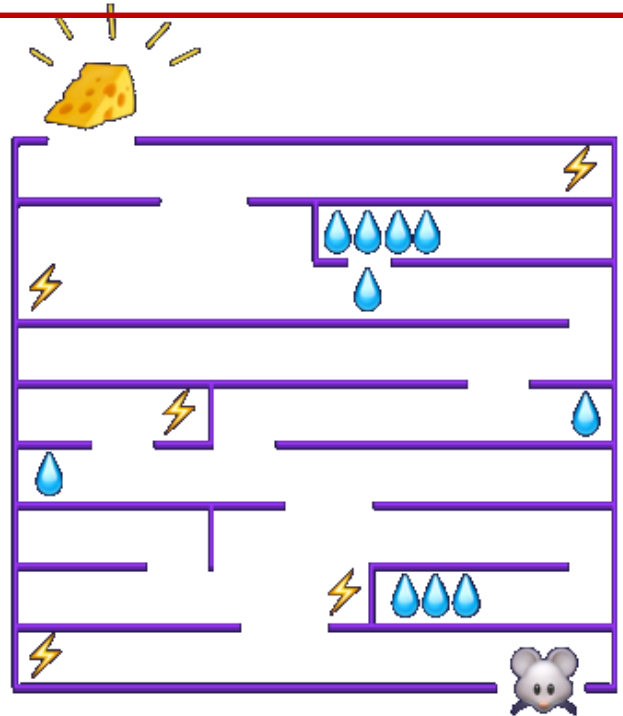
Query:
 $Q(\text{dòng}, \text{cột})$

$$Q(s^2, a^3) = 0.97$$

$$Q(s_t, a_t) \leftarrow (1 - \alpha) \cdot \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} \right)$$

- s_t : state s tại thời điểm t
- a_t : action a tại thời điểm t
- $Q(s_t, a_t)$: expected reward
- γ : Hệ số giảm ($0 \leq \gamma \leq 1$)

→ Cập nhật Q-table mỗi action

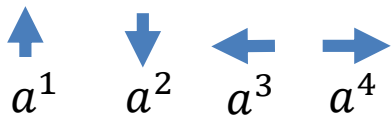


Chiến lược đi:

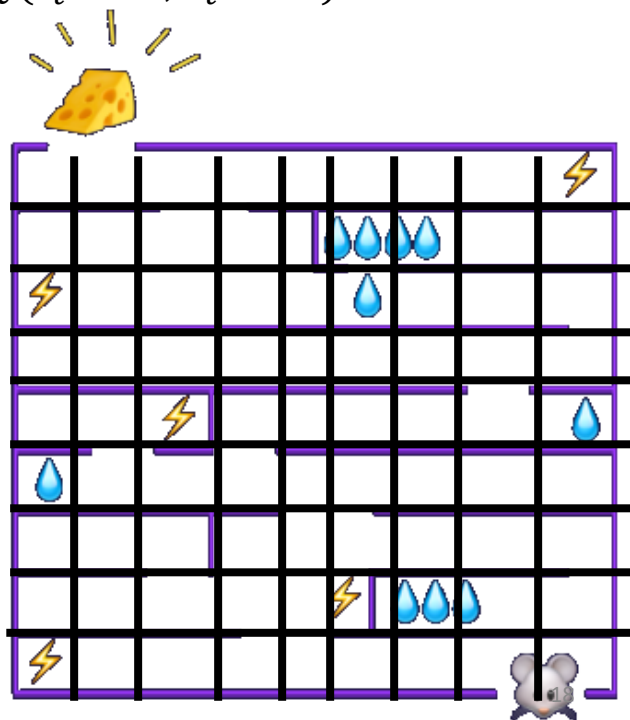
– Dựa vào giá trị lớn nhất trong Q-table.

VD khi $state = s^1$: $Q(s_t = s^1)$ thì có $Q(s_t = s^1, a_t = a^2)$ lớn nhất

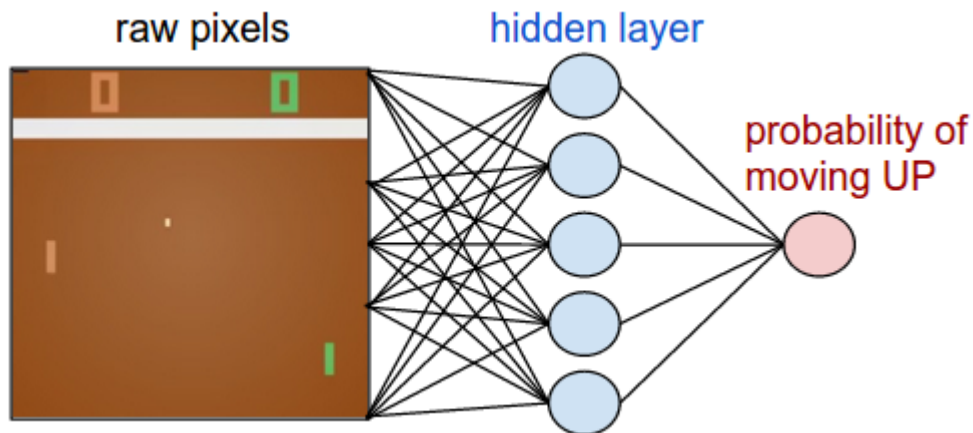
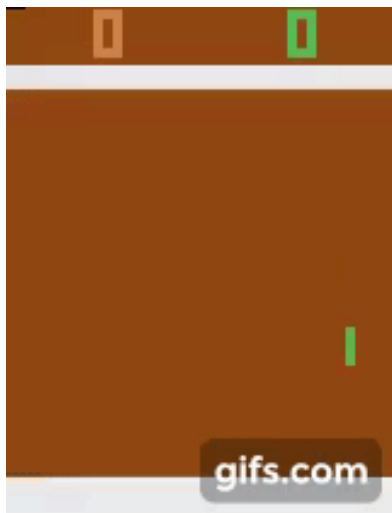
– Đưa vào softmax rồi lấy xí ngẫu

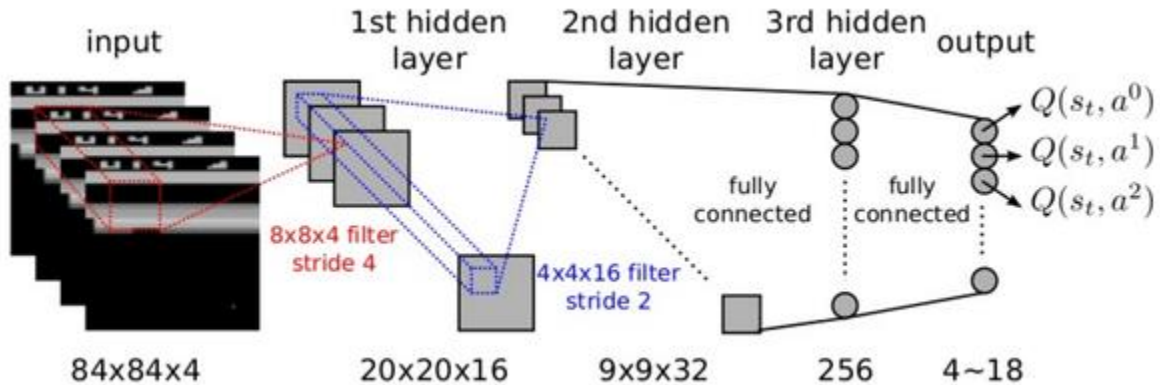


	a^1	a^2	a^3	a^4
s^1	0.23	0.81	0.4	0.54
s^2	0.1	0.54	0.97	0.2
s^3	0.3	0.81	0.76	0.3
...
s^n	0.23	0.81	0.4	0.54



- Cải tiến Q-learning sử dụng neural network
→ **Policy learning**
 - Policy learning dùng để map từ state sang action
- VD: Gặp đèn vàng → Giảm tốc độ





Layer	Input	Filter size	Stride	Num filters	Activation	Output
conv1	$84 \times 84 \times 4$	8×8	4	32	ReLU	$20 \times 20 \times 32$
conv2	$20 \times 20 \times 32$	4×4	2	64	ReLU	$9 \times 9 \times 64$
conv3	$9 \times 9 \times 64$	3×3	1	64	ReLU	$7 \times 7 \times 64$
fc4	$7 \times 7 \times 64$			512	ReLU	512
fc5	512			18	Linear	18

No pooling

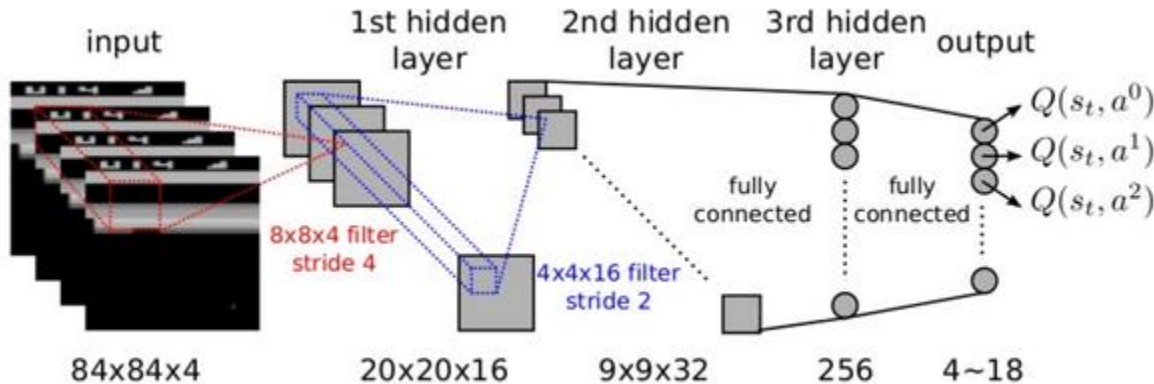
- Loss function của Deep Q network:

$$Loss = \frac{1}{2} \cdot \underbrace{\left[r + \max_{a_{t+1}} (Q(s_{t+1}, a_{t+1}; \theta_t)) \right]}_{\text{target}} - \underbrace{Q(s, a; \theta)}_{\text{prediction}}]^2$$

- Các bước thực hiện:**

1. Feedforward 1 lần để lấy đầy đủ giá trị $Q(s_t, a_t)$
2. Feedforward lần nữa để lấy đầy đủ giá trị $Q(s_{t+1}, a_{t+1})$
3. Lấy giá trị lớn nhất ở bước 2 gán cho target của action đó
4. Dùng back-propagation để học

$$Loss = \frac{1}{2} \cdot \underbrace{\left[r + \max_{a_{t+1}} (Q(s_{t+1}, a_{t+1}; \theta_t)) \right]}_{\text{target}} - \underbrace{Q(s, a; \theta)}_{\text{prediction}}]^2$$



$$Q(s_t) = [1.3, 0.4, \mathbf{4.3}, 1.5]$$

$$Q(s_{t+1}) = [\mathbf{9.1}, 2.4, 0.1, 0.5]$$

Cho $r = 2$

$$Loss = \frac{1}{2} \cdot (11.1 - 4.3)^2$$

- Exploration và Exploitation:
 - Exploitation: Đưa ra quyết định tốt nhất với lượng kiến thức hiện tại
 - Exploration: Thu thập thêm thông tin
- Dùng ϵ -greedy: chọn action random theo xác suất ϵ cho trước

$$a_t = \begin{cases} a_t, & \text{với xác suất } 1 - \epsilon \\ \text{action bất kỳ,} & \text{với xác suất } \epsilon \end{cases}$$

- Local-Minima:
 - Trong quá trình huấn luyện sẽ gặp nhiều ảnh giống nhau – học lại replay của những episode cũ
- Giải quyết bằng cách trộn vào minibatch những hình ảnh bất kỳ



Nội dung

1. Giới thiệu về re-inforcement learning
2. Động lực sử dụng deep re-inforcement learning
3. Bài toán sử dụng deep re-inforcement learning
4. Sử dụng deep re-inforcement learning như thế nào
5. Một số công trình về deep re-inforcement learning



Công trình nghiên cứu

Khóa học:

- <http://rail.eecs.berkeley.edu/deeprlcourse/>
- <https://www.udacity.com/course/reinforcement-learning--ud600>

Blog:

- <https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-0-q-learning-with-tables-and-neural-networks-d195264329d0>

Toolkit:

- OpenAI Gym: <https://gym.openai.com/>