Reinforcement Learning

Amir Abbasi

Dec 2020,University of Guilan

About the presenter

Amir Abbasi
Reinforcement Learning Researcher

amir.abbasi.rose@gmail.com

I've worked on Computer Vision for a year and it's about six months that I'm working on Reinforcement Learning.

- **Published papers:**

  Amir Abbasi , Erfan Miahi , S. A. Mirroshandel

  Effect of deep transfer and multi-task learning on sperm abnormality detection

  Computers in Biology and Medicine(2020)

  https://www.sciencedirect.com/science/article/pii/S0010482520304522?via%3Dihub

- **Current papers:**

  generating sperm images using GAN

  stock prediction using RL

  Image captioning using RL

Outline

- Introduction
- Trial and error learning
- Law of effect
- Reinforcement Learning
- Machine Learning
- Elements of Reinforcement Learning
- Types of RL algorithms
- RL in use
- Simulators
- The best RL courses
- Conclusion
- References
- Q&A

Introduction

The idea that we learn by interacting with our environment is probably the first to occur to us when we think about the nature of learning.Learning from interaction is a fundamental idea underlying nearly all theories of learning and intelligence.In this presentation,we are going to discuss about one of Machine Learning approaches that describes how to learn through interacting with environment.This approach is called “Reinforcement Learning” and it is much more focused on goal-directed learning from interaction than are other approaches to machine learning.

Trial And Error learning

R. S. Woodworth

According to American psychologist R. S. Woodworth (1938) :

The idea of trial-and-error learning goes as far back as the 1850s to Alexander Bain's discussion of learning by "groping and experiment" and more explicitly to the British ethologist and psychologist Conway Lloyd Morgan's 1894 use of the term to describe his observations of animal behavior.

(Sutton & Barto,p.15)

Law of effect

Edward Thorndike

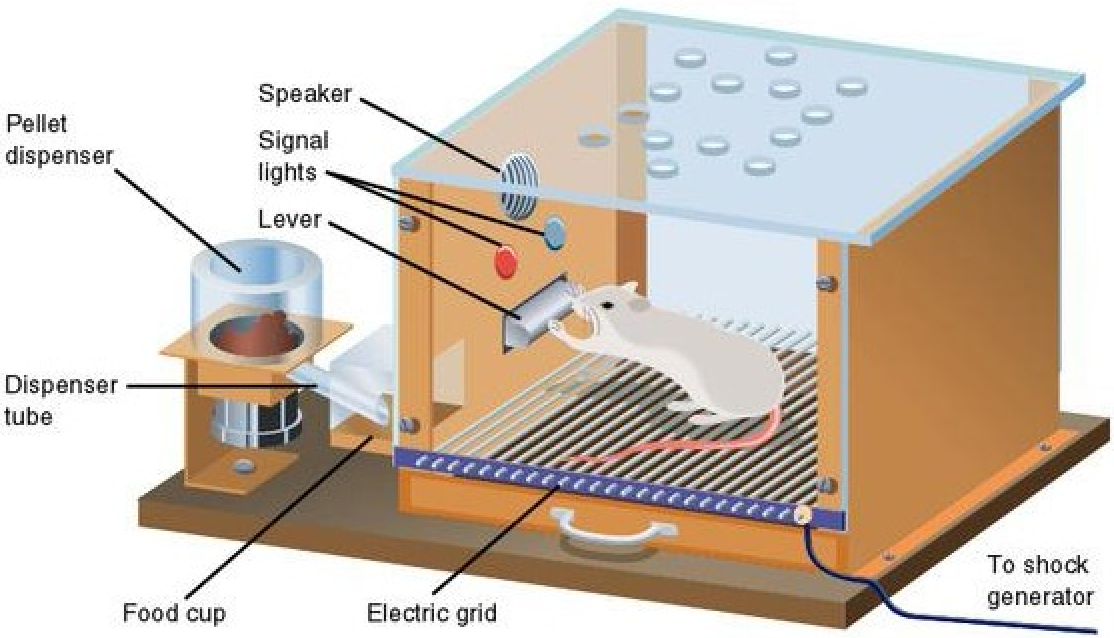
"Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs,they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond."

Thorndike, 1911, p. 244

(Sutton & Barto,p.15)

Law of effect(continue)

Responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation.



Operant conditioning chamber(skinner).

(Gray & Peter, p 108–109)

LOE is everywhere!

LOE and humunity

Envisage your childhood.How did you learn to walk?How did you learn to ride a bike?

Walking stages

A child learning riding a bike

LOE and humunity

We still learn some task by RL!

A man learning riding a motorcycle

Playing chess

Reinforcement Learning

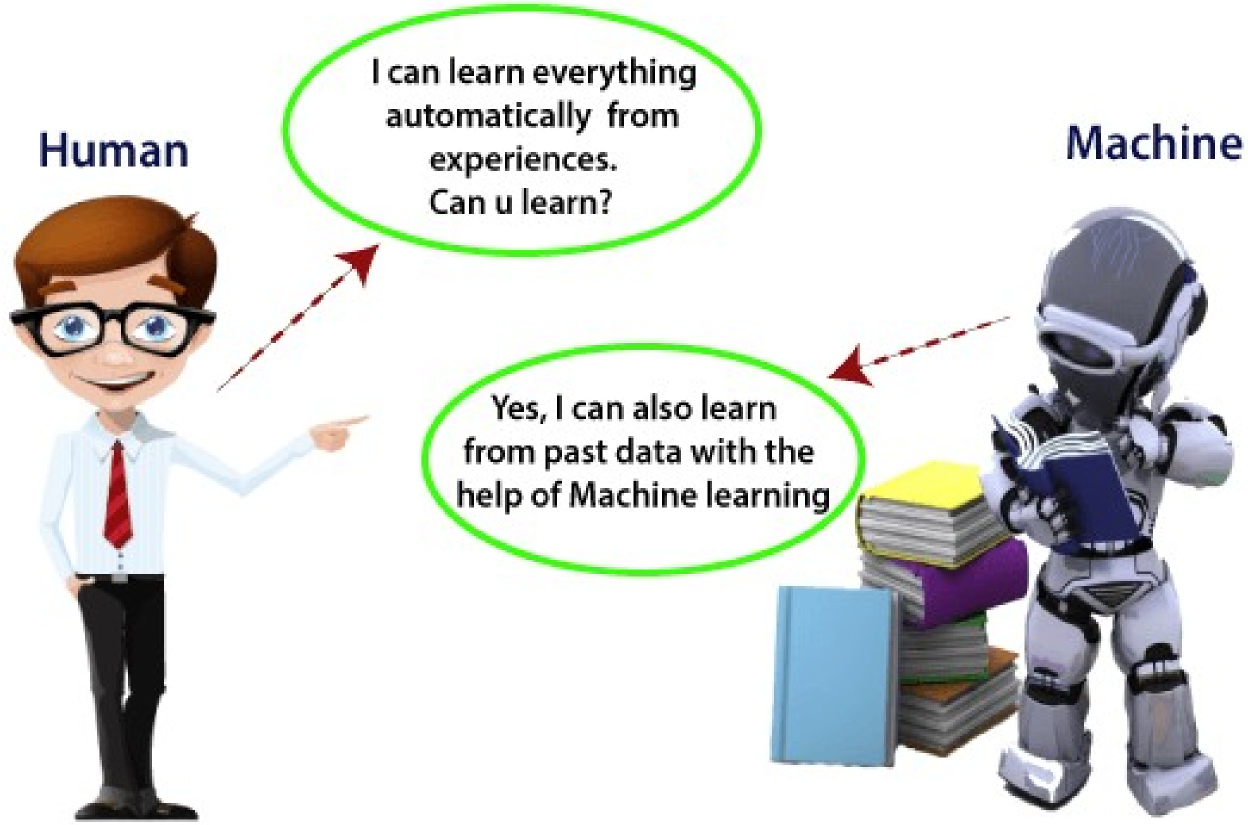
Let's utilize LOE on our computers to learn them what to do!

Richard Sutton(the father of RL)

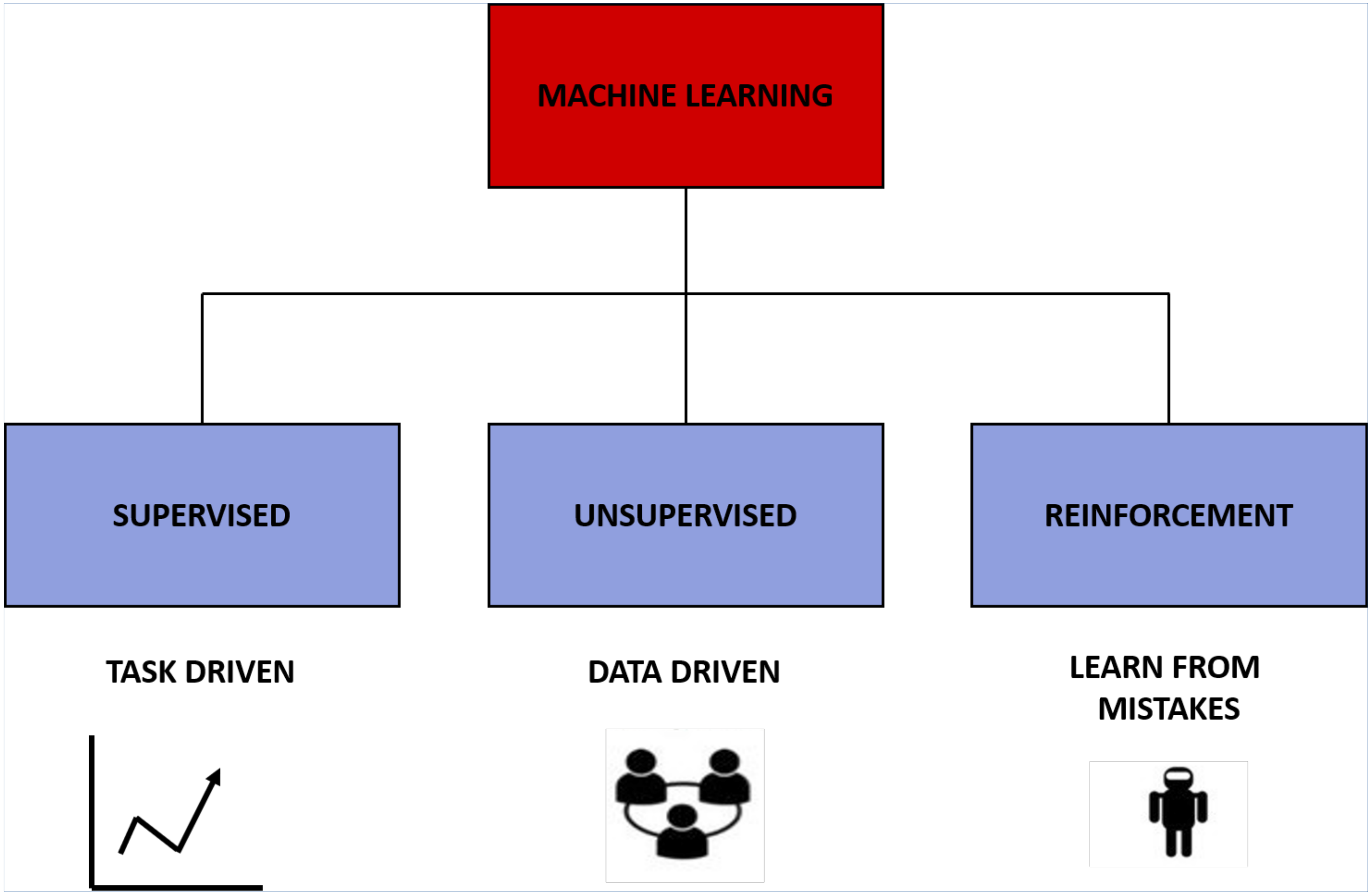
Andrew G. Barto

Machine Learning

Machine learning is the study of computer algorithms that improves automatically through experience.



Machine Learning



Machine learning comprises three branches

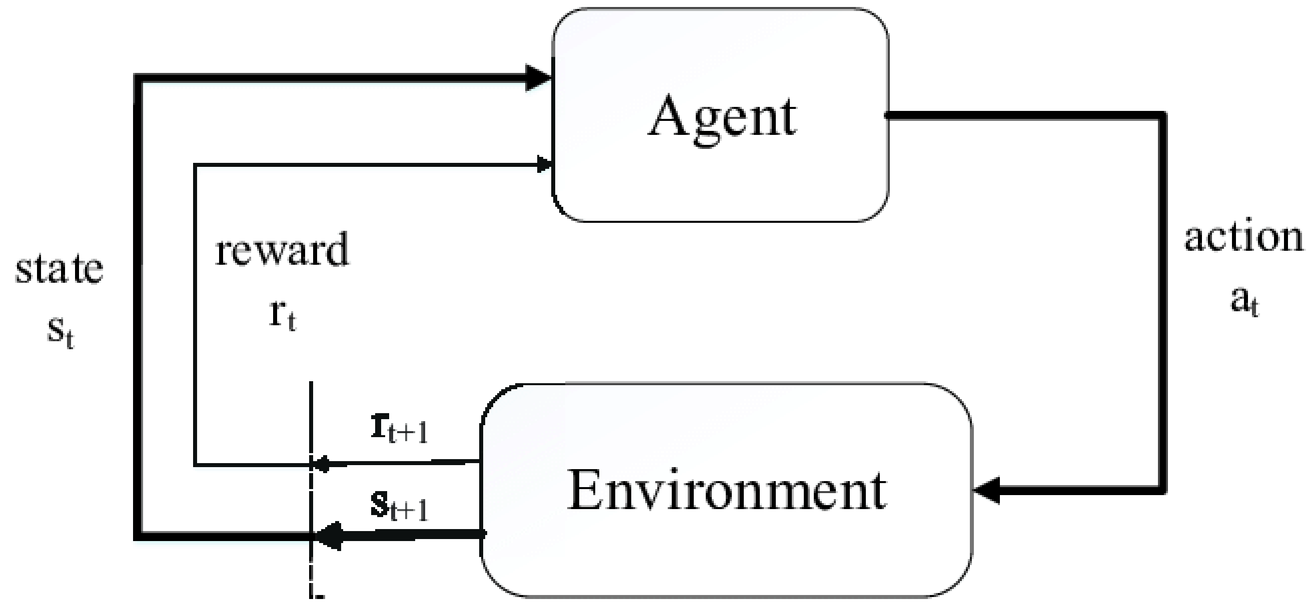
RL vs other methods

Basis for comparison	Supervised	Unsupervised	RL
Training data	Need domain expert to label data	Unlabeled data	Learn through interaction with environment
Preference	Routine tasks (input output mapping)	Clustering, discovering data correlation and new patterns	Artificial intelligence (Behavioral learning)
Area	Machine learning	Machine learning	Machine learning
Optimal strategy	Depend on the data and learning algorithm	Depend on the data and its classification	Learn optimal strategy from experience
Exploration	No exploration	No exploration	Adaptable to changes through exploration

Elements of Reinforcement Learning

Beyond the agent and the environment, one can identify four main subelements of a reinforcement learning system:

- policy
- reward signal
- value function
- model of the environment(optional)



Main elements in RL

(Sutton & Barto,p.6)

Reward

“That all of what we mean by goals and purposes can be well thought of as the maximization of the expected value of the cumulative sum of a received scalar signal (called reward).”

Suttion & Barto

$$r(s,a,s') \doteq \mathbb{E}[R_t \mid S_{t-1}=s, A_{t-1}=a, S_t=s'] = \sum_{r\in\mathcal{R}} r \frac{p(s',r \mid s,a)}{p(s' \mid s,a)}$$

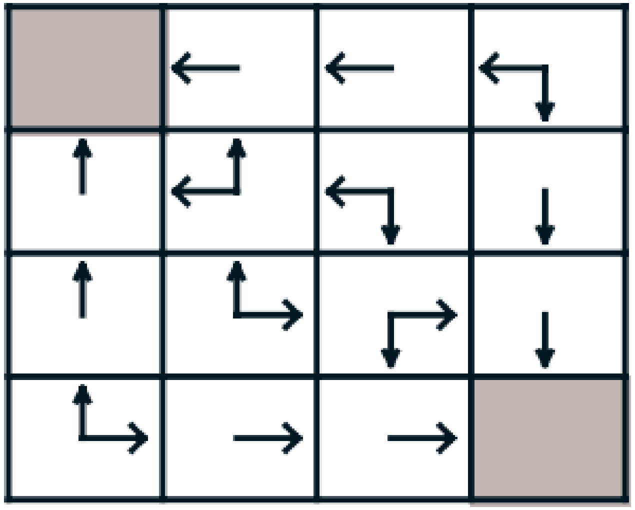
In fact,if the sequence of rewards received after time step t is denoted R_{t+1}, R_{t+2}, R_{t+3}, . . .,the goals is to maximize expected return:

$$G_t \doteq R_{t+1} + R_{t+2} + R_{t+3} + \cdots + R_T$$

(Sutton & Barto,p. 48-49 & 53)

Policy

A policy is a mapping from states to probabilities of selecting each possible action and is show as $\pi(a|s)$.In fact,$\pi(a|s)$is the probability that $A_t = a$ if $S_t = s$.



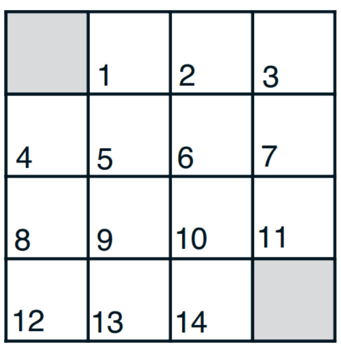
A grid world that has two terminal states

(Sutton & Barto,p.58)

Value function

The value function of a state s under a policy π, denoted $v_\pi(s)$, is the expected return when starting in s and following π thereafter.

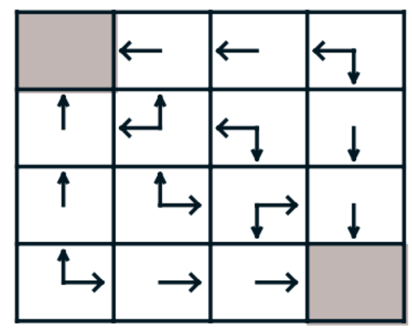
$$v_\pi(s) \doteq \mathbb{E}_\pi[G_t \mid S_t = s]$$



A grid world

0.0	-14.	-20.	-22.
-14.	-18.	-20.	-20.
-20.	-20.	-18.	-14.
-22.	-20.	-14.	0.0

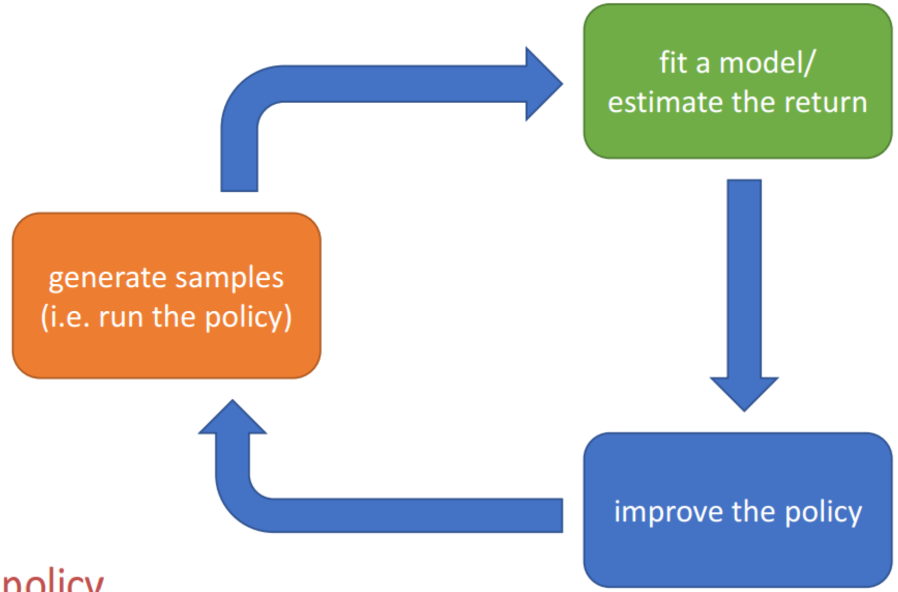
Value


Policy

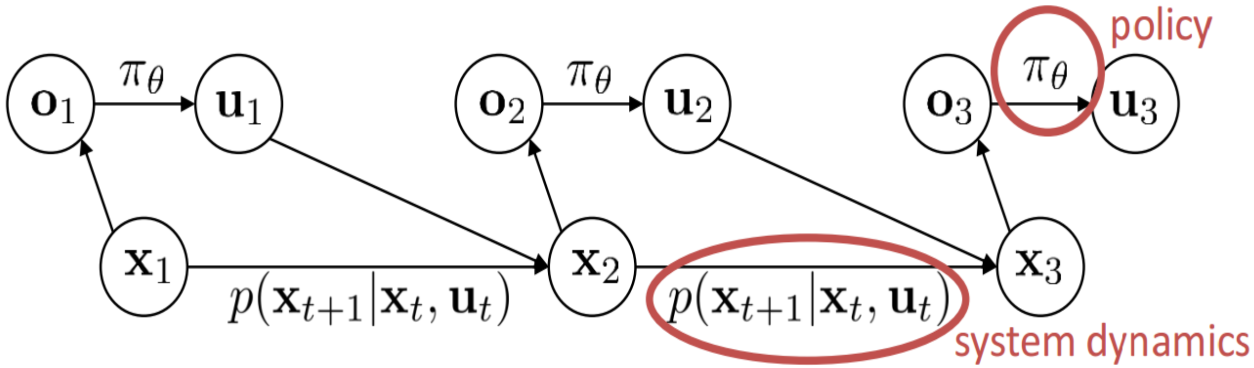
(Sutton & Barto,p.58)

Types of RL algorithms

In RL,there are three main steps:
1. Run the policy
2. Fit a model/estimate the return
3. Improve the policy



Main steps in RL algorithms.



We can choose next state using neither dynamic or π .

(Levine's slides)

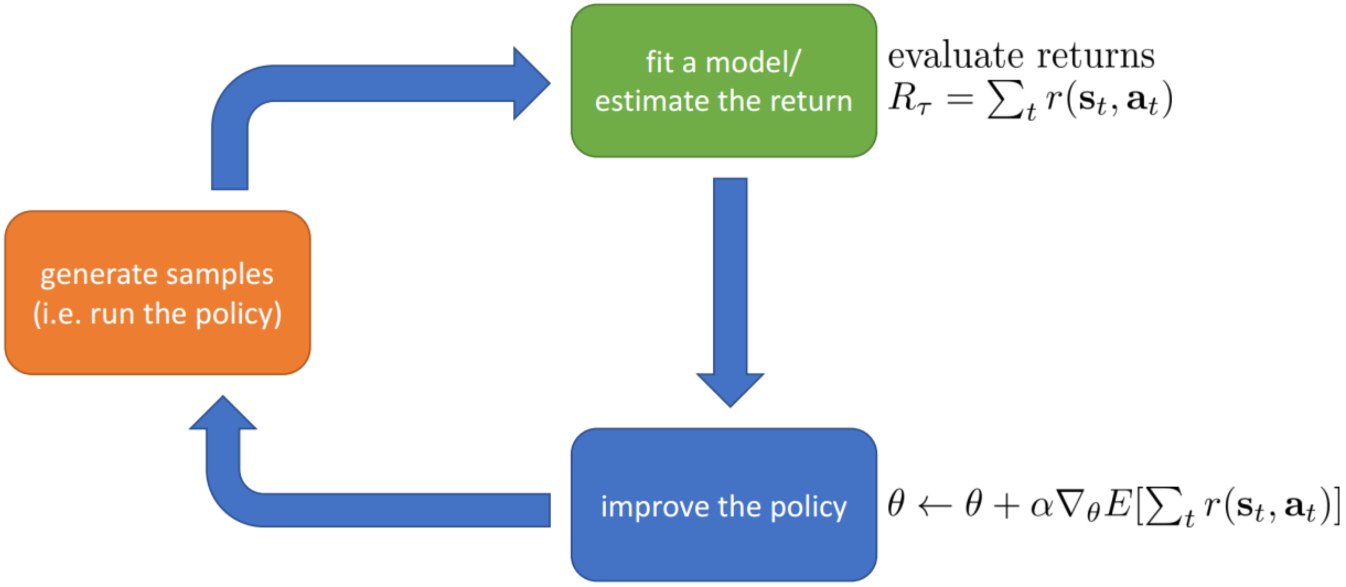
Types of RL algorithms

- Policy gradients
- Value-based
- Actor-critic
- Model-based RL: estimate the transition model, and then…
    1. Use it for planning (no explicit policy)
    2. Use it to improve a policy

(Levine's slides)

Policy gradient

In policy gradient methods,optimal policy is learned by computing gradient of policy and then optimizing it.

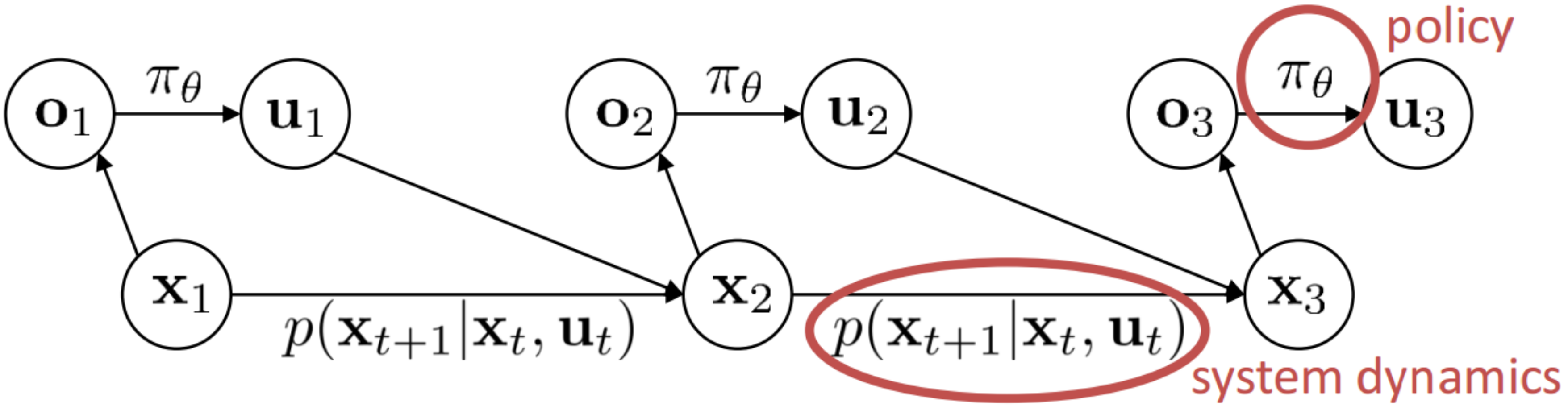


Policy gradient

(Levine’s slides)

Model-based

In model-based methods,the goal is to use known dynamics or learn dynamics.



In model-based settings,we want to learn system dynamics or use them to plan.

(Levine's slides)

Value based

In value based methods,value function or Q-function of the optimal policy is estimated.

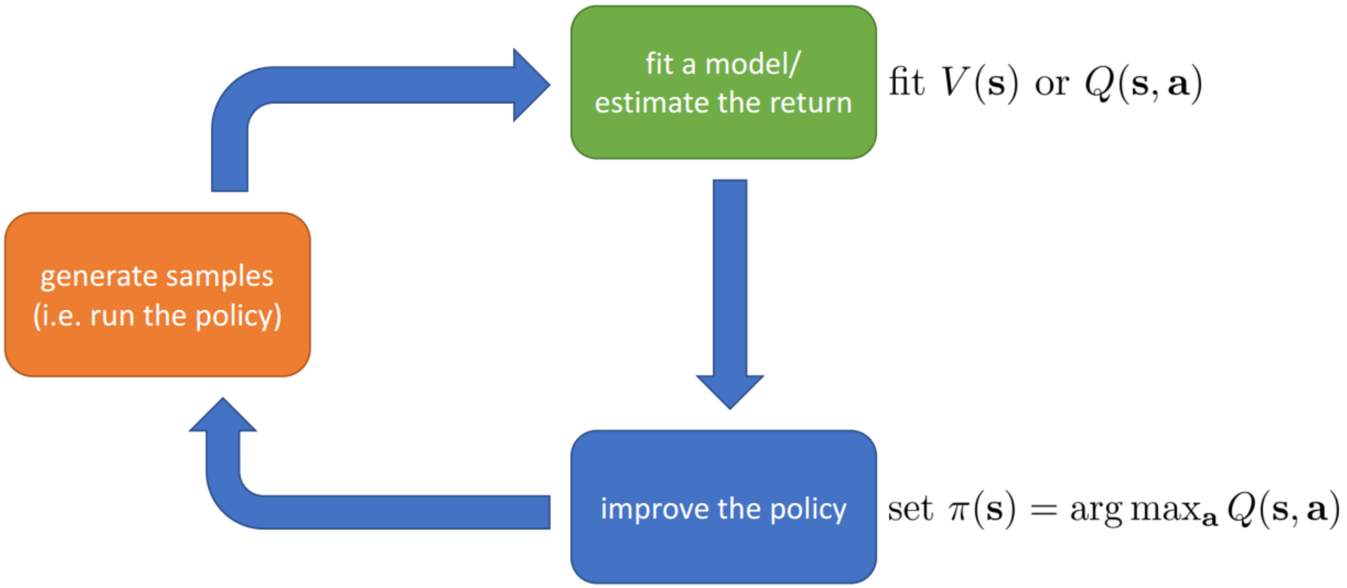
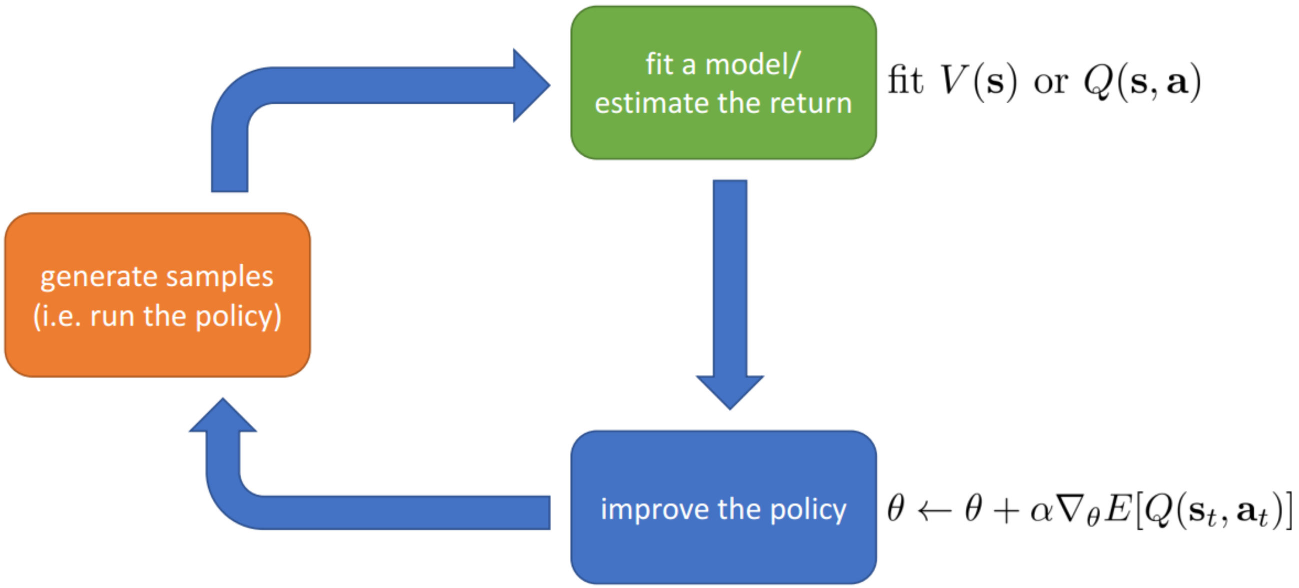


Figure 19:Value based

(Levine's slides)

Actor-Critic

In Actor-critic methods,value function or Q-function of the current policy is estimated and used to improve policy.



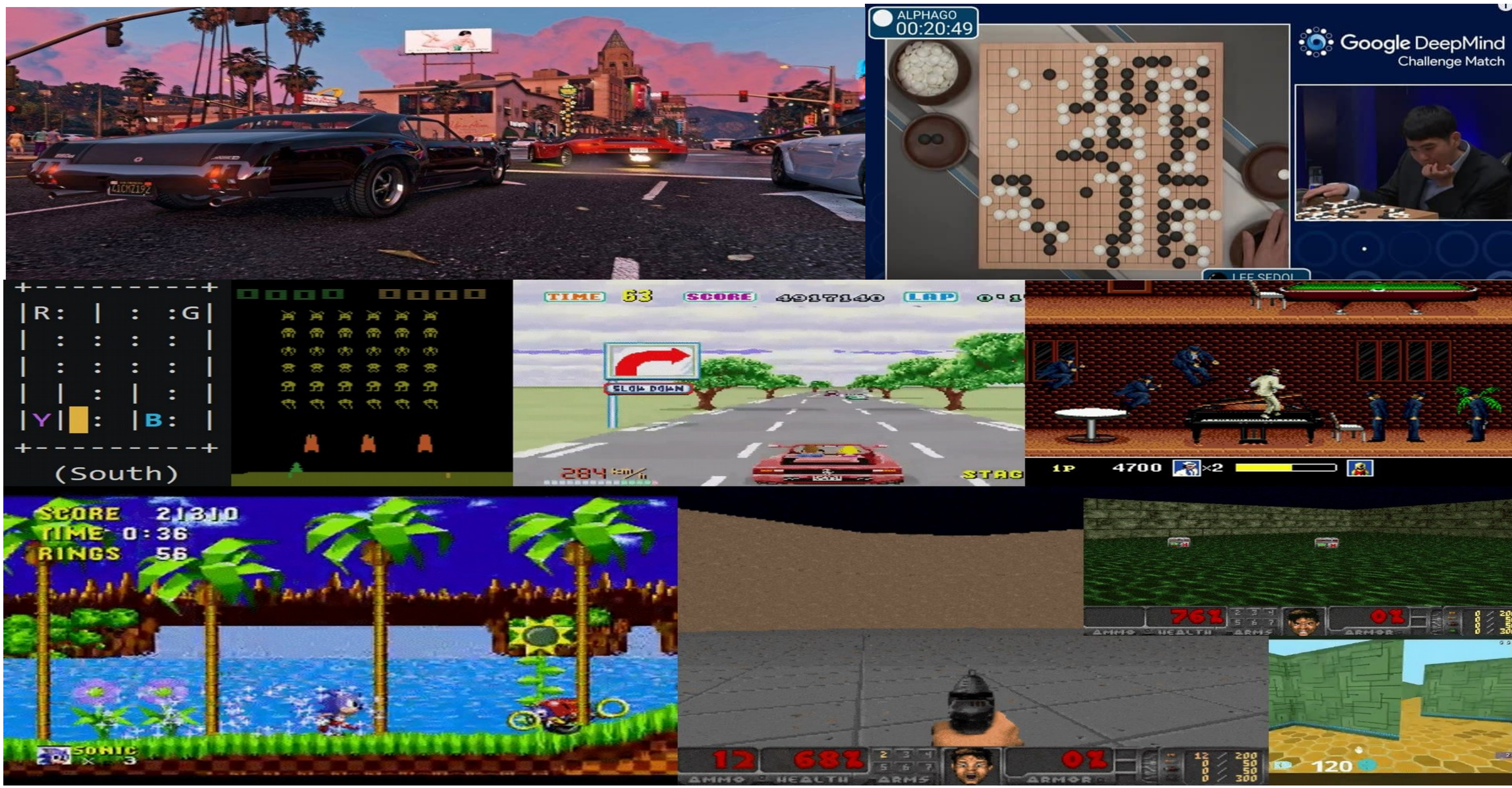
Actor-critic

(Levine's slides)

RL in use

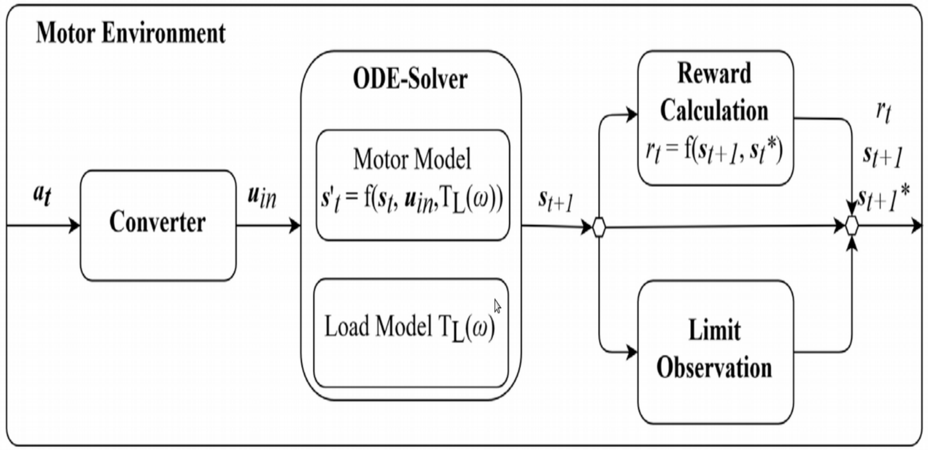
- Games
- Electrical engineering
- Business Process Management
- Autonomous Driving
- Robotics
- Cyber security
- Other Machine Learning fields such as Computer Vision,NLP and etc.
- ...

RL and Games



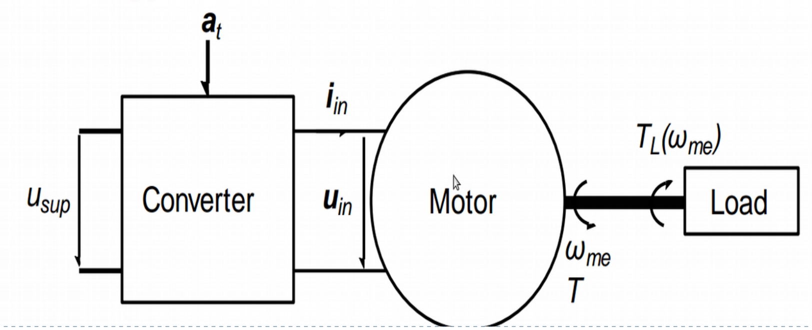
ALPHAGO
00:20:49
Google DeepMind
Challenge Match
R:
:G
Y
B:
(South)
TIME 63
SCORE 4917140
LAP
SLOW DOWN
1P 4700 x2
SCORE 21310
TIME 0:36
RINGS 56
SONIC x 3
12
68%
0%
AMMO
HEALTH
ARMS
ARMOR
120

Electrical engineering



Environments for simulation of electric motors in Python

gym-electric-motor (GEM) toolbox



1 of 1

دانشگاه صنعتی شریف

دانشکده مهندسی برق

وبینار علمی با موضوع :

تخصیص منابع در مخابرات با استفاده از یادگیری تقویتی

سخنران : دکتر مهدی نوری بروجردی (محقق پسا دکتری)

زمان: شنبه ۹۹/۹/۲۲ – ساعت: ۱۷:۱۵ لغایت ۱۸

آدرس لینک وبینار : https://vc.sharif.edu/ch/ee-webinar2

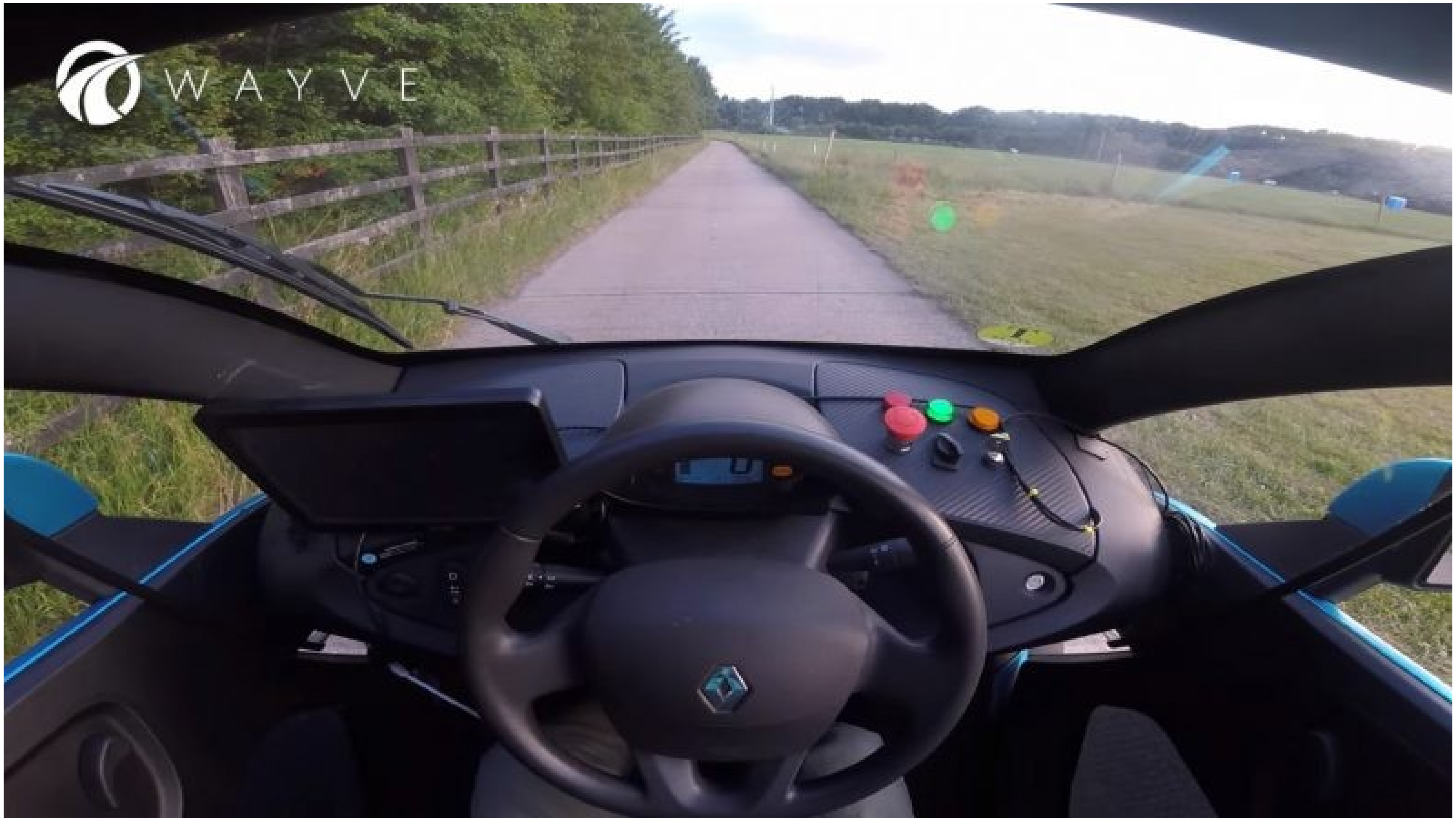
Deep Reinforcement Learning for Power System Applications: An Overview

Zidong Zhang, Dongxia Zhang, and Robert C. Qiu, *Fellow, IEEE*

Abstract—Due to increasing complexity, uncertainty and data dimensions in power systems, conventional methods often meet bottlenecks when attempting to solve decision and control problems. Therefore, data-driven methods toward solving such problems are being extensively studied. Deep reinforcement learning (DRL) is one of these data-driven methods and is regarded as real artificial intelligence (AI). DRL is a combination of deep learning (DL) and reinforcement learning (RL). This field of research has been applied to solve a wide range of complex sequential decision-making problems, including those in power systems. This paper firstly reviews the basic ideas, models, algorithms and techniques of DRL. Applications in power systems such as energy management, demand response, electricity market, operational control, and others are then considered. In addition, recent advances in DRL including the combination of RL with other classical methods, and the prospect and challenges of applications in power systems are also discussed.

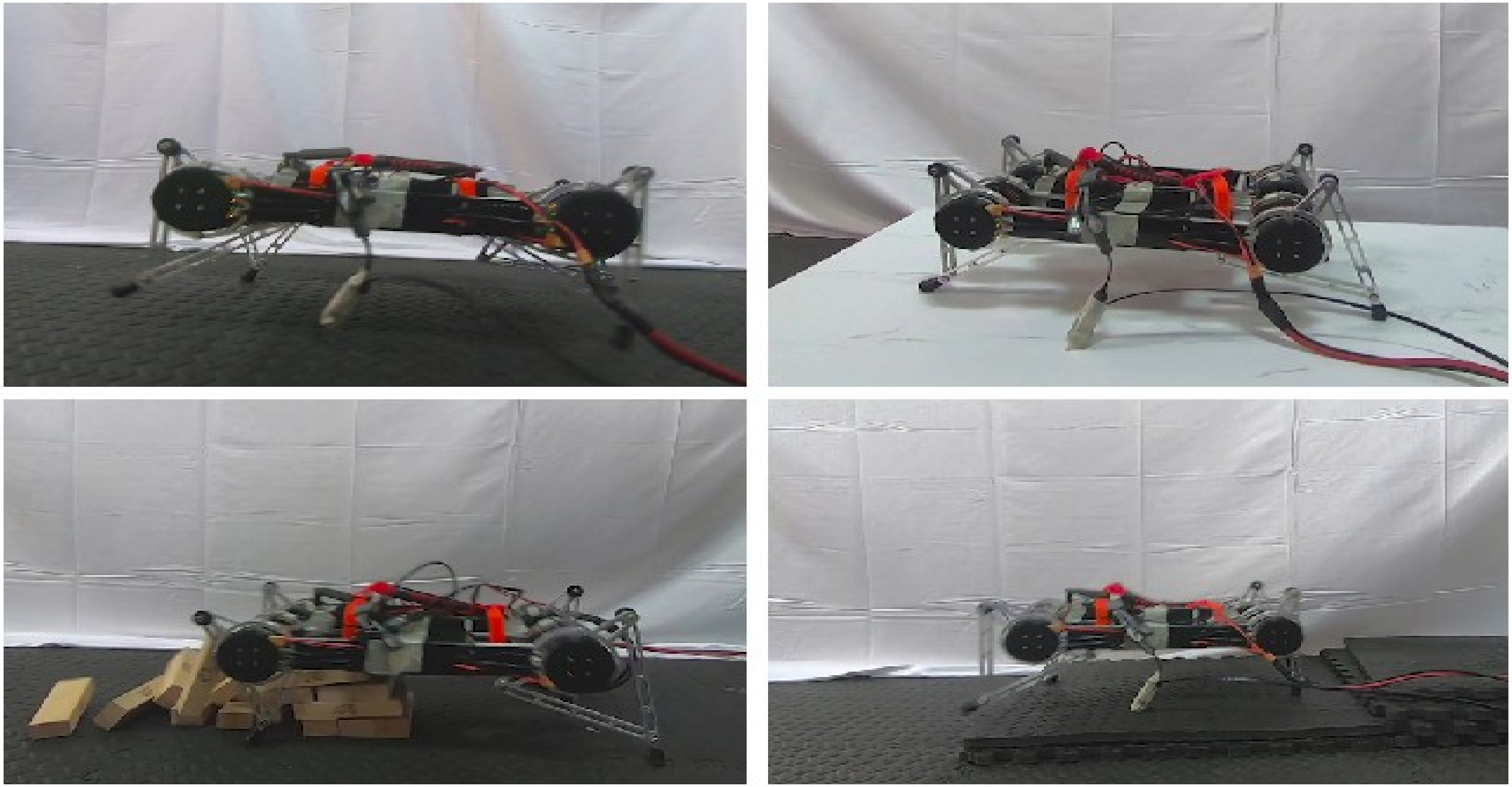
DFRL	Deepforest reinforcement learning.
DG	Distributed generation.
DL	Deep learning.
DNN	Deep neural network.
DPG	Deterministic policy gradient.
DQL	Deep Q-learning.
DQN	Deep Q-network.
DR	Demand response
DRL	Deep reinforcement learning.
DTQ	Deep transfer network.
EH	Energy harvesting.
EI	Energy Internet.
EM	Electricity market.
EMA	Exponential moving average.
ES	Energy storage.

Autonomous driving



A self-driving car that is invented by Wayve Corporation.

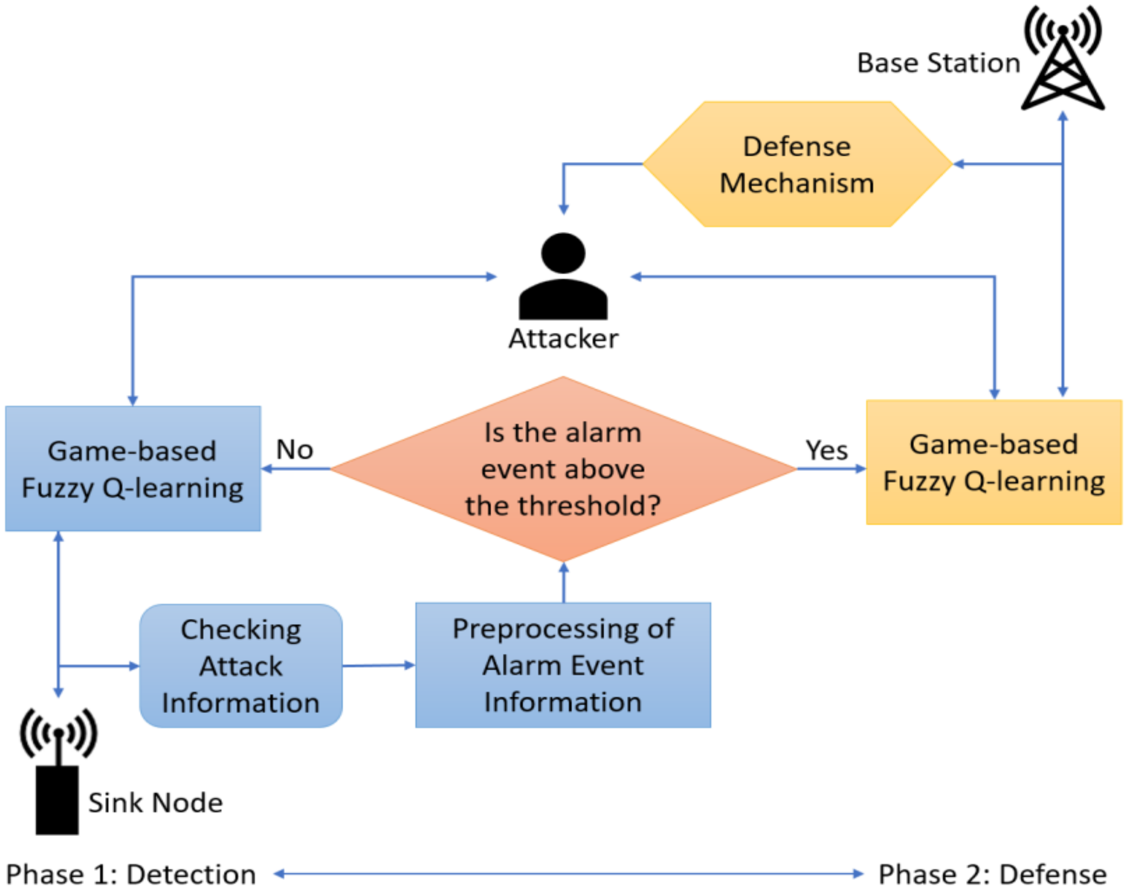
Robotics



A robot learning to walk using Deep RL.

(Haarnoja,Ha, Zhou,Tan, Tucker,.2020)

Cyber security

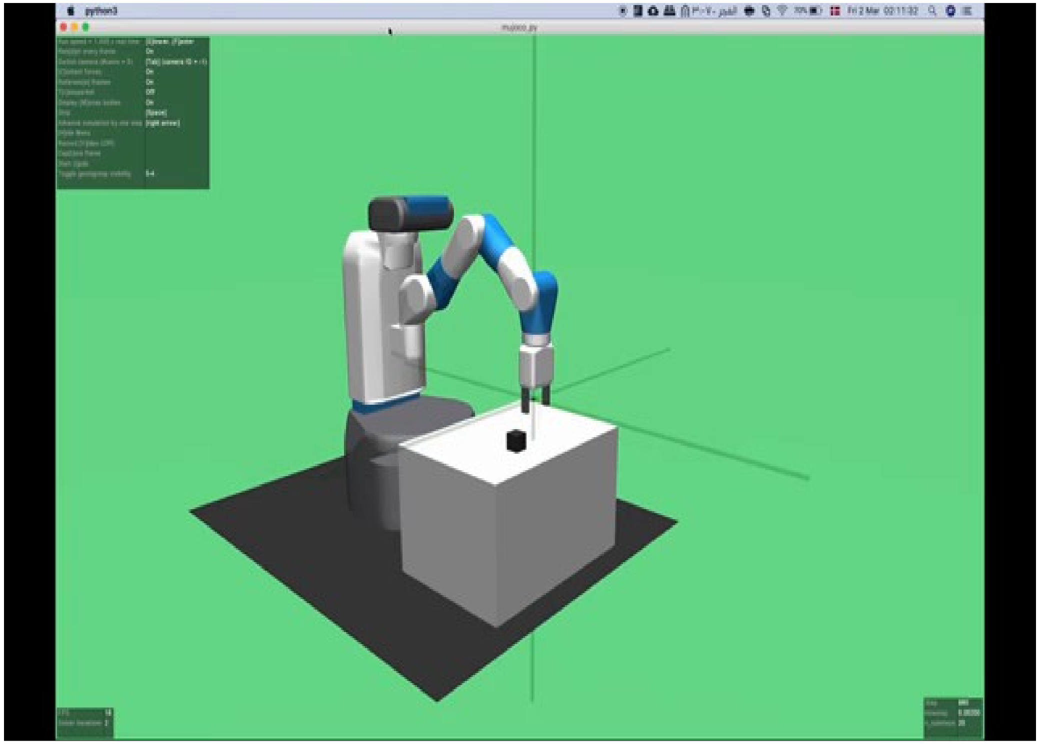


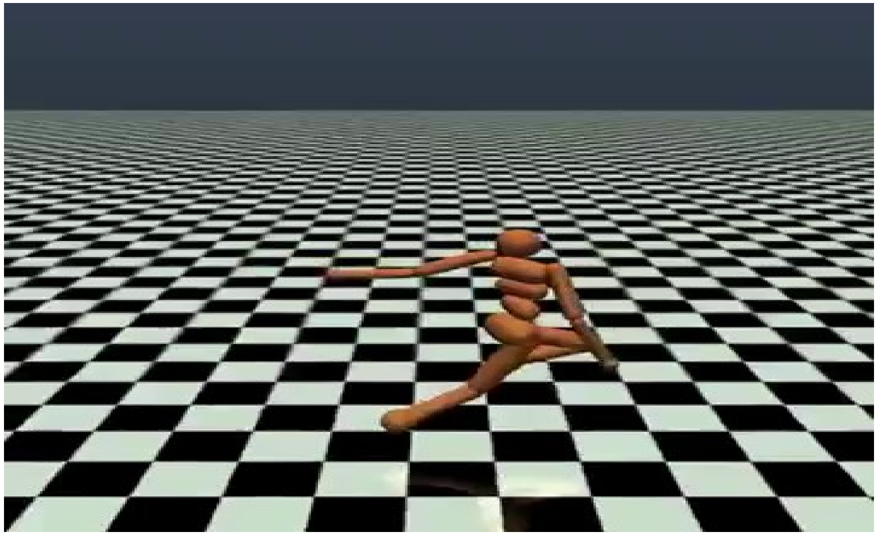
Intrusion detection and prevention system based on a game theory approach and fuzzy Q-learning

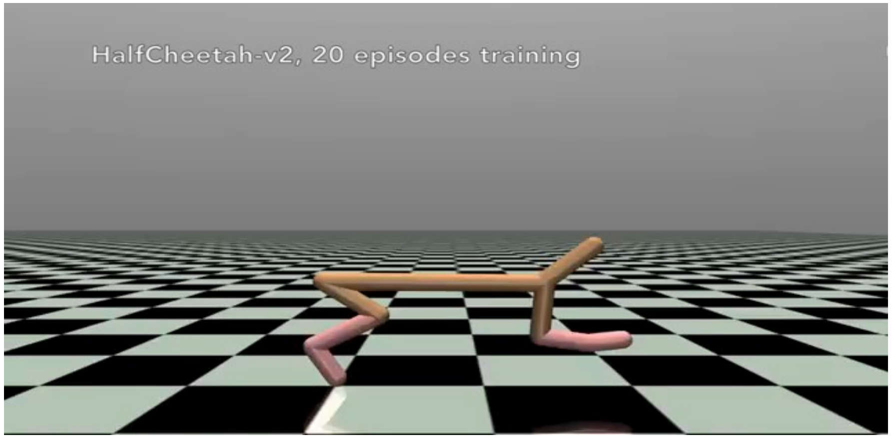
(Nguyen,Reddi.2020)

Simulators

- Some of gym simulators:







The best RL courses

Reinforcement Learning Specialization

UNIVERSITY OF ALBERTA

Prospective Students

Home | Admissions & Programs | Faculties | Research | Campus Life | News | About U of A | COVID-19 Information

/ Admissions & Programs / Online Courses / Reinforcement Learning

Reinforcement Learning

About the Specialization

The **Reinforcement Learning Specialization** consists of **4 courses** exploring the power of adaptive learning systems and artificial intelligence (AI).

Harnessing the full potential of AI requires adaptive learning systems; this is exactly what reinforcement learning (RL) does by design: improve through trial-and-error interaction.

By the end of this Specialization, learners will understand the foundations of much of modern probabilistic AI and be prepared to take more advanced courses, or to apply AI tools and ideas to real-world problems. This content will focus on "small-scale" problems in order to understand the foundations of Reinforcement Learning.

The tools learned in this Specialization can be applied to:

- AI in game development,
- IOT devices,
- Clinical decision making,
- Industrial process control,
- Finance portfolio balancing,
- & more.

RL

https://www.coursera.org/learn/fundamentals-of-reinforcement-learning

Deep Reinforcement Learning

CS 285 at UC Berkeley

Deep Reinforcement Learning

Lectures: Mon/Wed 5:30-7 p.m., Online

Lectures will be recorded and provided before the lecture slot. The lecture slot will consist of discussions on the course content covered in the lecture videos.

Piazza is the preferred platform to communicate with the instructors. However, if for some reason you wish to contact the course staff by email, use the following email address: cs285fall2020@googlegroups.com.

Lecture recordings from the current (Fall 2020) offering of the course: watch here
Enrolled students: please use the private link you were provided, not this one!

Looking for deep RL course materials from past years?
Recordings of lectures from fall 2019 are here, and materials from previous offerings are here.

Instructor **Sergey Levine**
svlevine@eecs.berkeley.edu
Office Hours: After lecture

Head GSI **Michael Janner**
janner@eecs.berkeley.edu
Office Hours: Tuesday 3-4 pm

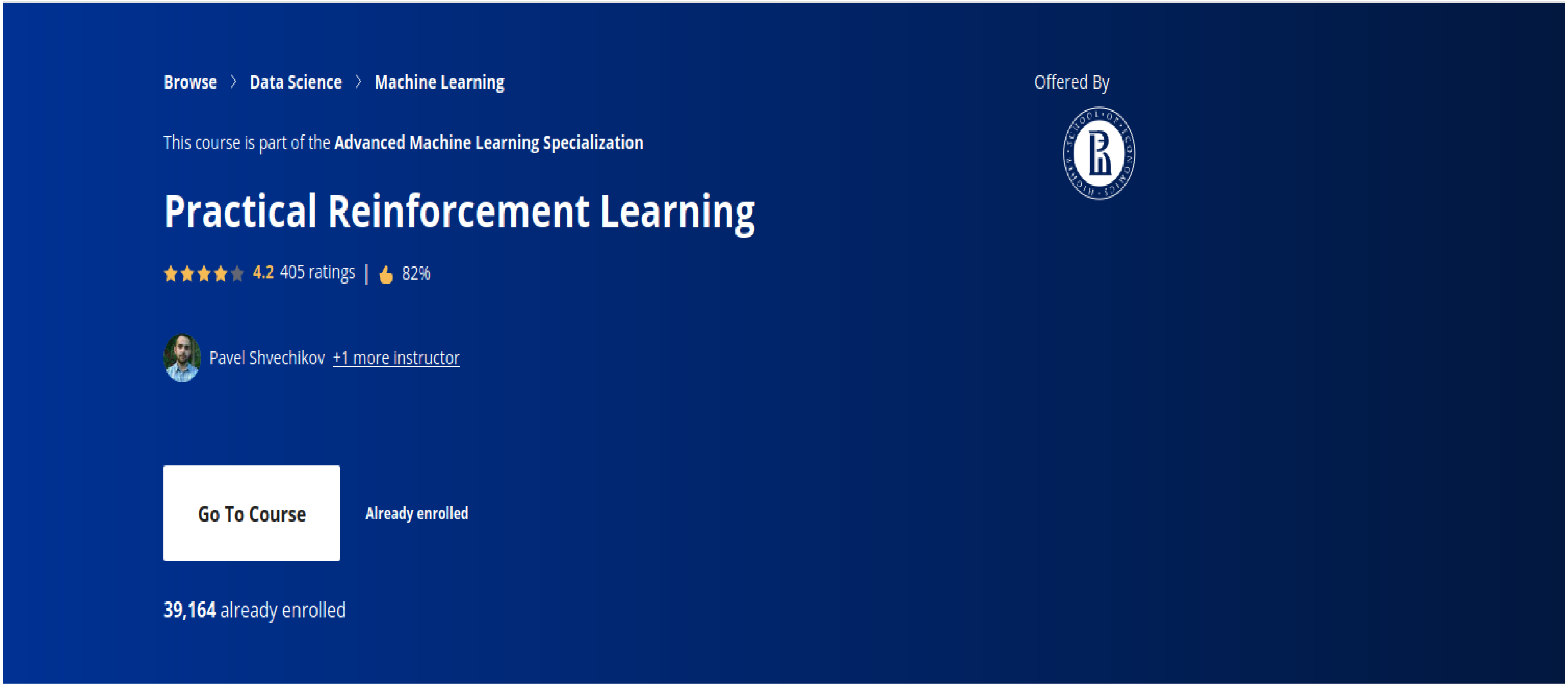
GSI **Vitchyr Pong**
vitchyr@eecs.berkeley.edu
Office Hours: Friday 1-2pm

GSI **Aviral Kumar**
aviralk@berkeley.edu
Office Hours: Thursday 2-3pm

uGSI **Alexander Khazatsky**
khazatsky@berkeley.edu
Office Hours: Monday 12-1pm

http://rail.eecs.berkeley.edu/deeprlcourse/

Prcatical Reinforcement Learning



https://www.coursera.org/learn/practical-rl

Conclusion

In this presentation,we talked about Fundamentals of Reinforcement Learning and attempted to illustrate some of its major usages.As been discussed,RL come in handy not only in a wide range of AI tasks,but also in some other engineering fields such as Motor Control,Robotics,Self-driving and etc.

We hope that RL finds a good answer for some of tasks that aren’t solved yet.

References

1.Richard S. Sutton and Andrew G. Barto.Reinforcement Learning,An Introduction.second edtition.

2.Slides of CS 285,UC Berkeley.

3.Gray, Peter. Psychology, Worth, NY. 6th ed. Pp 108–109.

4.Official web site of Wayve corporation.

5.Haarnoja, Tuomas & Zhou, Aurick & Ha, Sehoon & Tan, Jie & Tucker, George & Levine, Sergey. (2018). Learning to Walk via Deep Reinforcement Learning.

6. Nguyen,Reddi.(2020).Deep Reinforcement Learning for Cyber Security.

Thanks for your attention!
Any questions?