Multi-Agent Reinforcement Learning

Yuxuan XIE

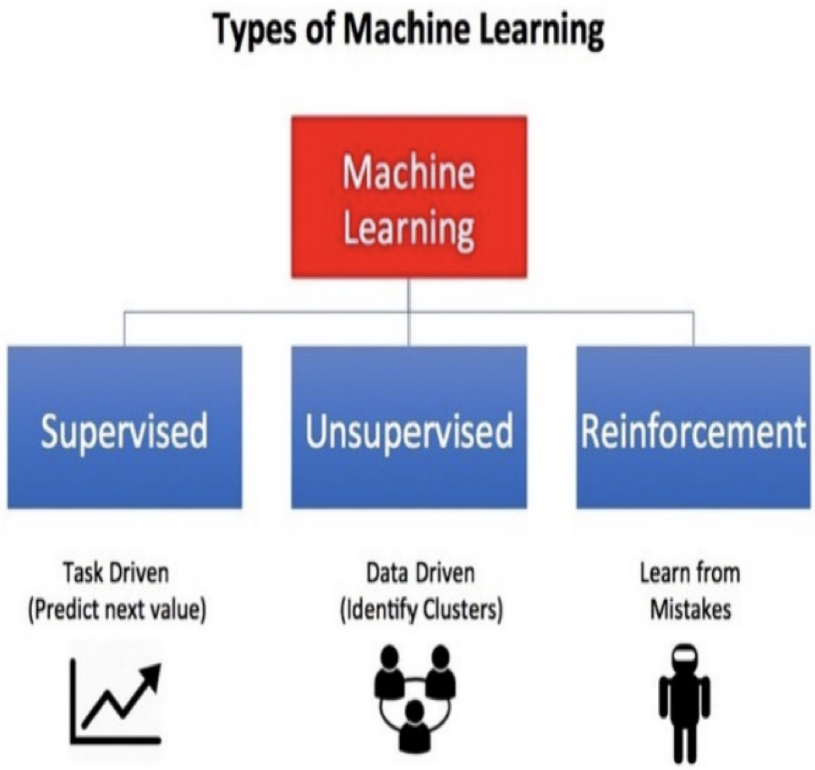
yuxuan.xie@hotmail.com

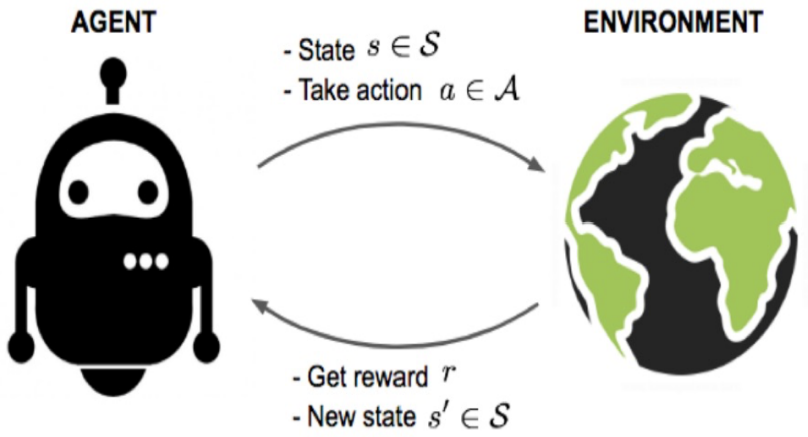
October 26, 2020

Outline

1. Single Agent Reinforcement Learning
2. Multi Agent RL and its challenges
3. Methods to solve but not optimally
4. Applications

The overview of RL





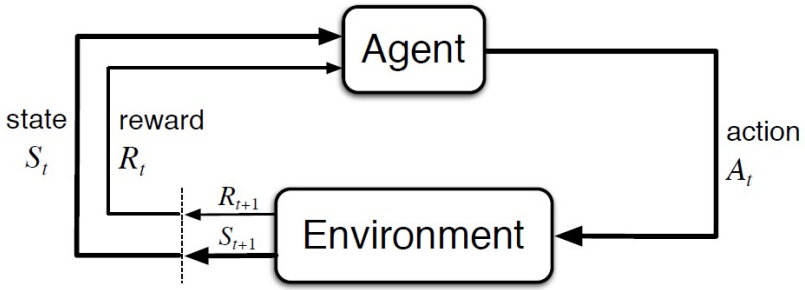
Learning from interacting with the environment

Why RL is so hot?

DL + RL = AGI

—David Silver

Markov Decision Process(MDP)



The MDP can be represented as a four element tuple $< S, A, R, P >$

- S : the set of state
- A : the set of action
- R : the set of reward
- P : the transition model

model-based and model-free

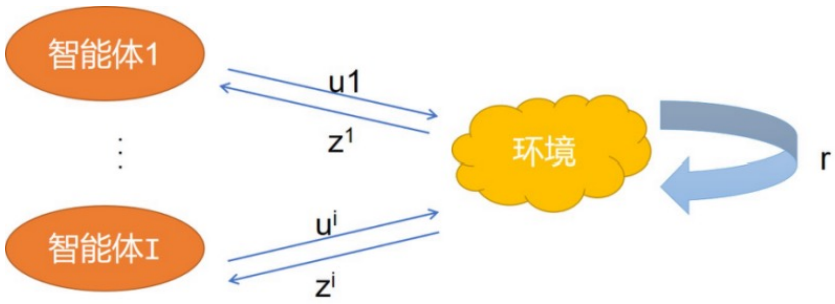
Methods to solve model-based RL

- Dynamic Programming
- Dyna(learn a model)

Methods to solve model-free RL

- Tabular based temporal-Difference Learning, e.g. Q-learning, SARSA
- Neural network based TD learning, e.g. DQN, DDQN
- Policy Gradient e.g. Actor-Critic, A3C, TRPO, PPO

Dec-POMDPs



A tuple $< S, A, R, P, O, N >$

- S : the set of state
- A : the set of action $A = \times_i A^i$
- R : the set of reward
- P : the transition model
- O : the set of observation $O = \times_i O^i$
- N : the number of agent

Model-based and Model-free

Methods to solve model-based MARL

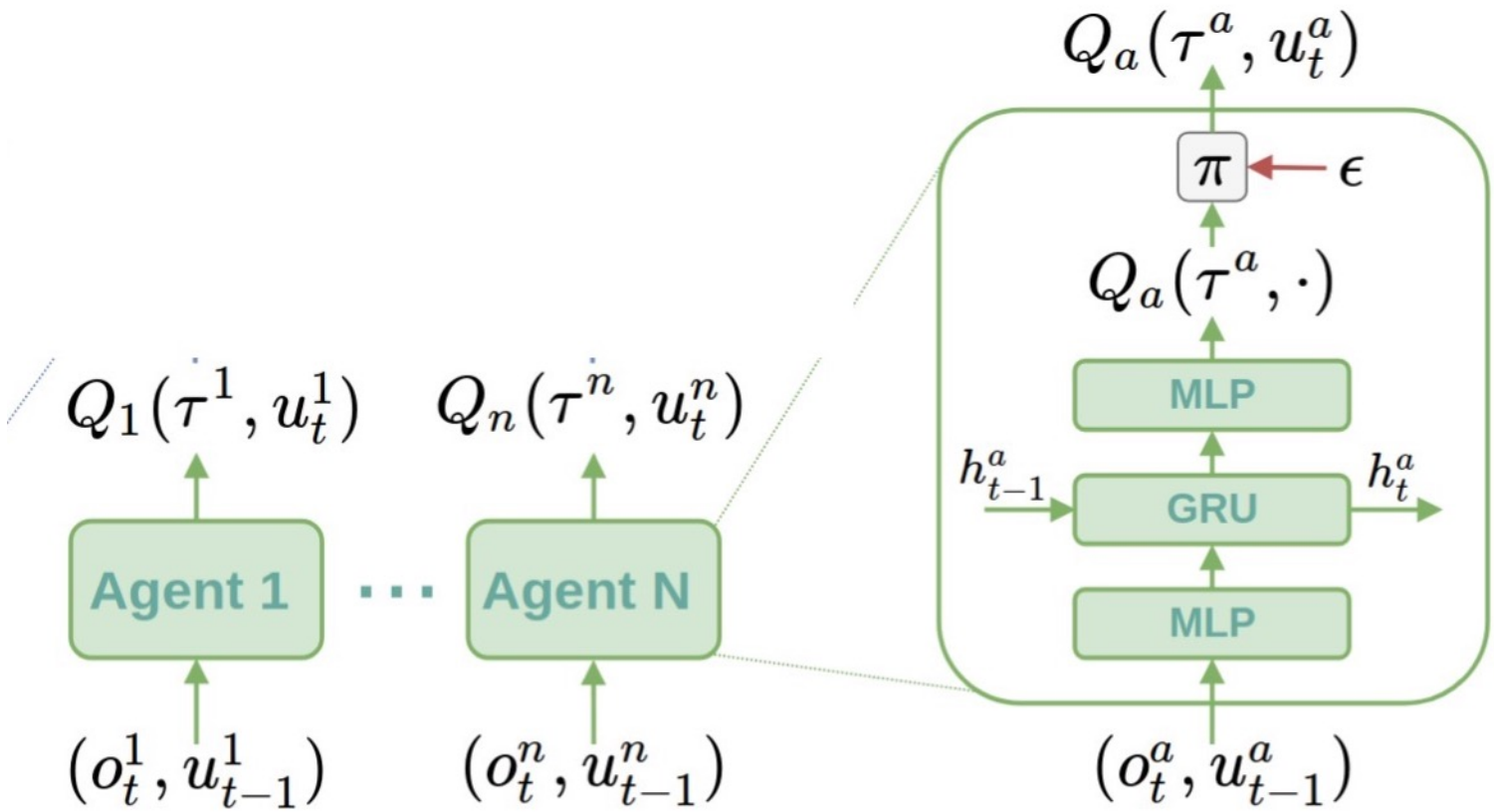
- Counterfactural regret minimazition (CFR)
- Occupancy-belief-state-HSVI

Methods to solve model-free MARL

core mechanism : Centralizd trainning and decentralized execution

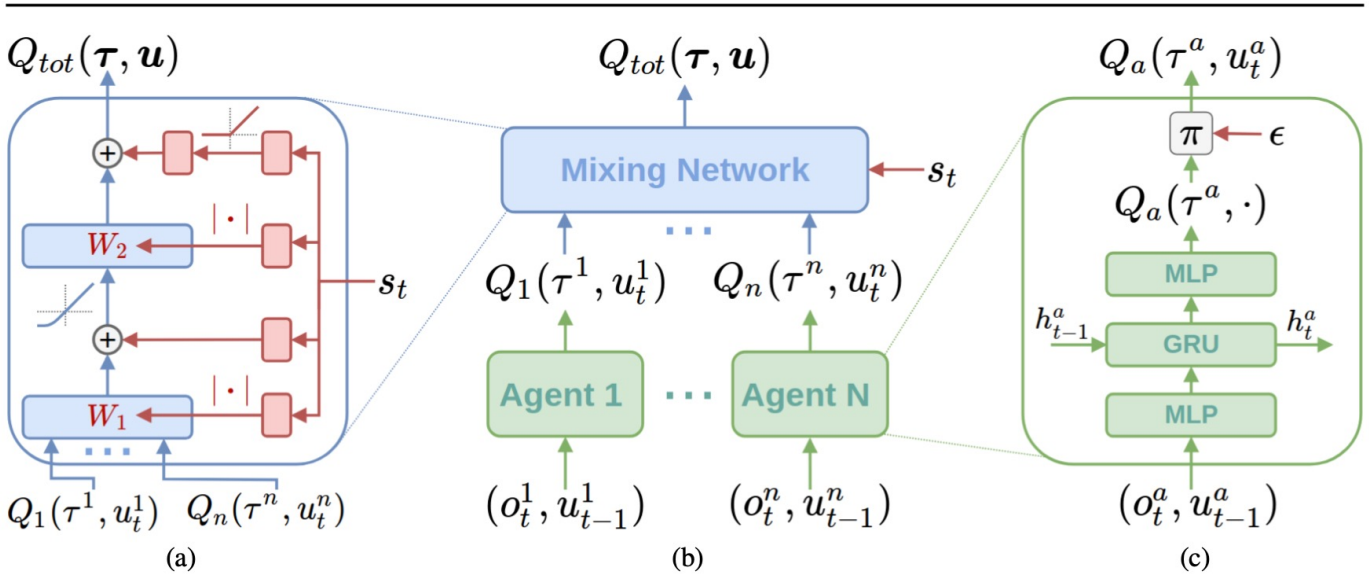
- Value based : IQL, VDN, QMIX, QTRAN,
- Policy based : COMA, MADDPG

Independent Q-Learning



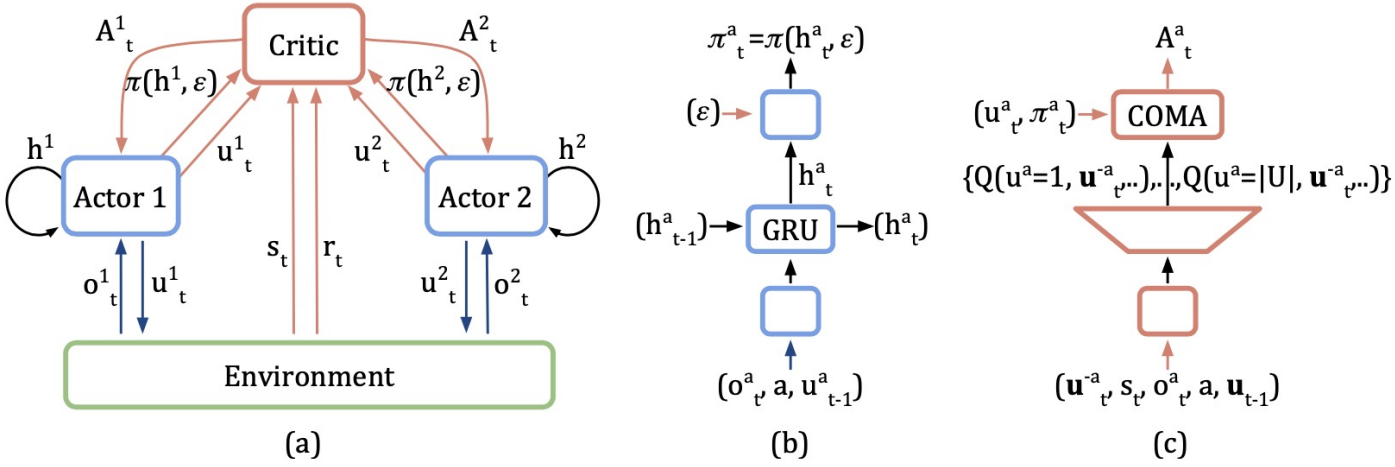
Every agent makes their own decision and others are treated as the environment. Then the environment is dynamic.

VDN and QMIX

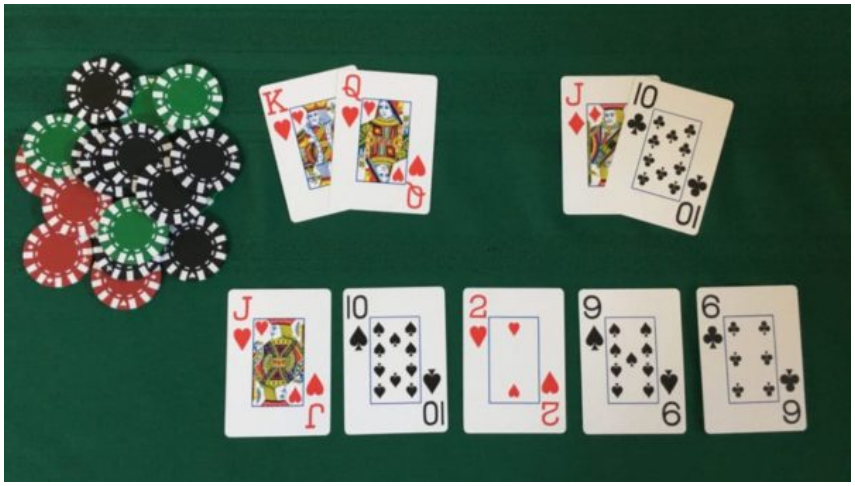


The mix network just works during the trainning process.
For VDN its a sum operator.

COMA



Just replace the value-based actor by the policy gradient actor.
And some changes about the policy gradient algorithms.



Teaxs hold'em

Hanabi

Starcraft

dota