Multi-agent reinforcement learning

Schnebli Zoltán [1]

[1] Babeş-Bolyai University, Faculty of Mathematics and Computer Science

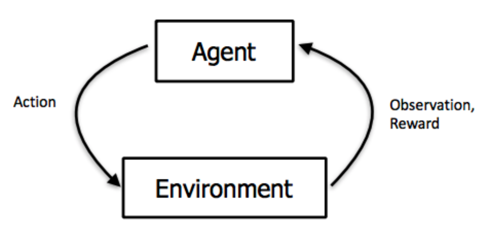
Introduction

- Motivation?
  - Automatization
  - Robotics

Reinforcement learning

Working principle:

- Agent
- Environment
- Action - State
- Reward



Reinforcement learning

Exploration vs. exploitation

- ϵ - greedy strategy
- ϵ - decay

Single agent environment

Markov decision process

- $\langle S, A, \mathcal{P}_{\cdot}(\cdot,\cdot), \mathcal{R}_{\cdot}(\cdot,\cdot), \gamma \rangle$
  - S - set of states
  - A - set of actions
  - $\mathcal{P}_a(s, s')$ - probability of reaching state s'
  - $\mathcal{R}_a(s, s')$ - value of the reward if we go to s'
  - γ - discount factor

Single agent environment

Partially observable Markov decision process

- $\langle S, A, P_{\cdot}(\cdot,\cdot), R_{\cdot}(\cdot,\cdot), \gamma, \Omega, O(\cdot,\cdot)\rangle$
  - S, A, T, R, γ
  - Ω - set of all observations
  - $\mathcal{O}_a(o, s')$ - probability of getting observation o

Multi agent environment

Markov games

- N agents
    - $\mathcal{A} := \{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n\}$
    - $\mathcal{O} := \{\mathcal{O}_1, \mathcal{O}_2, ..., \mathcal{O}_n\}$
- It is the most general modell

Recent research results

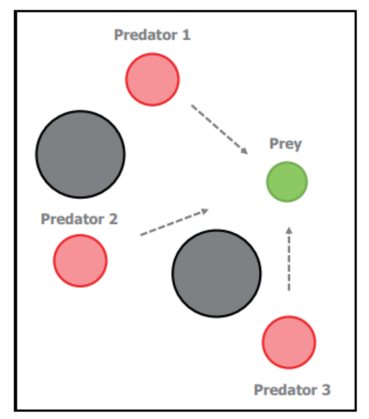
Deep deterministic policy gradient algorithm with generative cooperative policy network

- Every agent has 3 policies
  - Q-network -> optimal action during execution
  - Greedy policy network -> maximizes the global objective based on the local actions
  - Generative cooperative policy newtork -> learn other agents policies during training
- pro: cooperativeness
- con: extra policies to train

Recent research results

Experiment - Compared algorithms

- CF - shared
- FDMARL - individual
- DDPG
- DDPG-GCPN
- DDPG-GCPN with random GCPNs in sample-generating

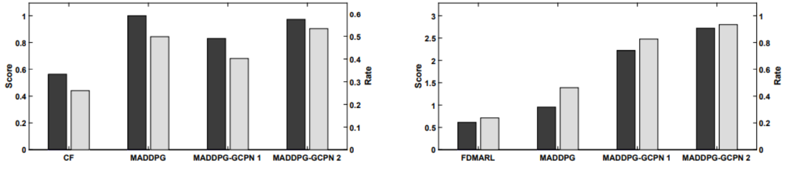


Recent research results

Experiment - Results

2 reward functions

- shared reward (a)
- individual (b)



(a)

(b)

Recent research results

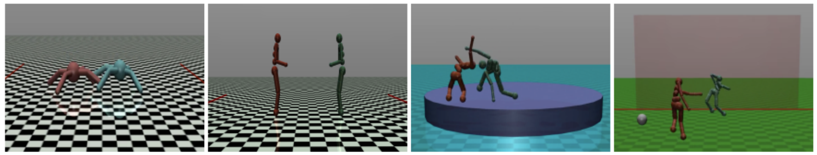
Emergent Complexity via Multi-Agent Competition

- goal: get complex agent behavior from simple environments
- ideea: self-play

Recent research results

Environments

- Run to Goal
- You Shall Not Pass
- Sumo
- Kick and Defend



Recent research results

Experiment - Results

- opponent sampling - random old opponent better
- exploration curriculum - dense reward at the beginning to learn basic motor skills faster
- interesting behaviors: blocking, rising arms, charging, kicking high, etc.

Thanks for watching