北京航空航天大學

A Survey on Large-Population Systems and Scalable Multi-Agent Reinforcement Learning

Kai Cui, Anam Tahir, Gizem Ekinci, Ahmed Elshamanhory, Yannick Eich, Mengguang Li and Heinz Koeppl

TECHNISCHE UNIVERSITÄT DARMSTADT

黄靖宜

2022/12/10

CONTENTS

INTRODUCTION

INTRODUCTION

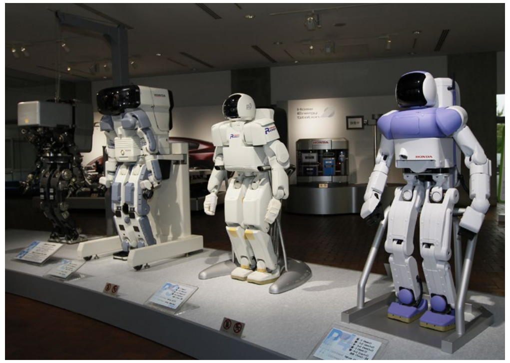
SEQUENTIAL DECISION-MAKING

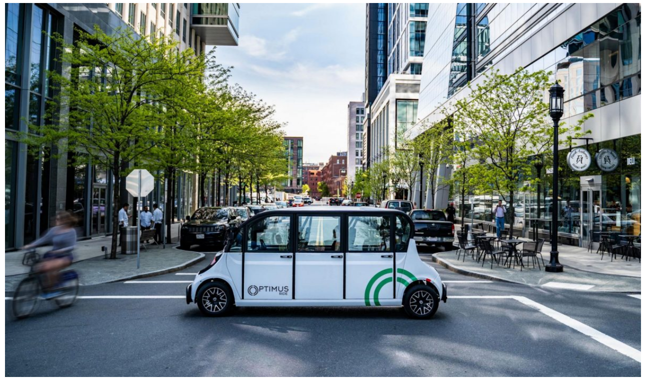
LARGE-POPULATION SYSTEMS

APPLICATIONS

FUTURE DIRECTIONS


video games


robotic systems


autonomous vehicles

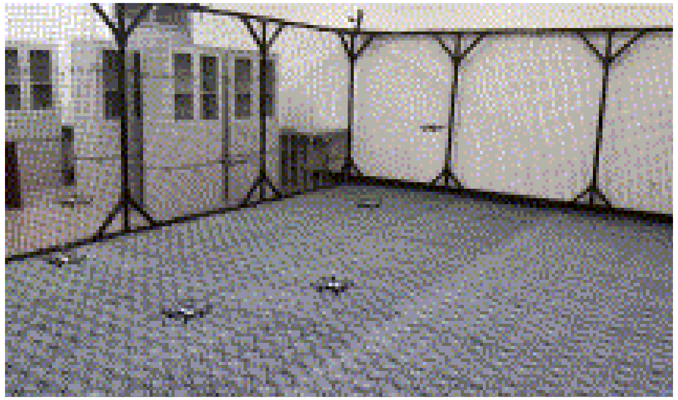

finance systems

Multi-Agent Reinforcement Learning

Multi-Agent Reinforcement Learning difficulties :

- nonuniqueness of learning goals
- non-stationarity of other learning agents
- **scalability to large state and action spaces of large numbers of agents**

→ the curse of many agents



multi-agent problems

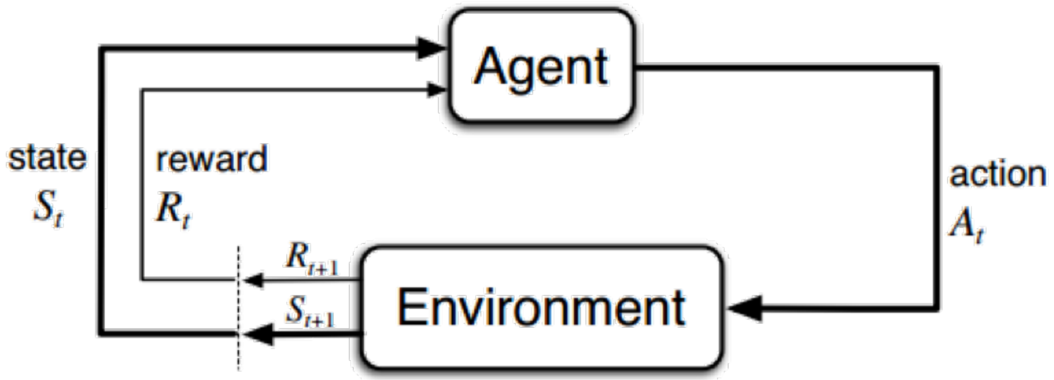
large-population problems

CONTENTS

➢ Single-agent reinforcement learning

Markov decision process :

a state space, an action space, transition function, and a reward function



Value-based :

- Q-learning
- DQN
- DDQN

Policy-based :

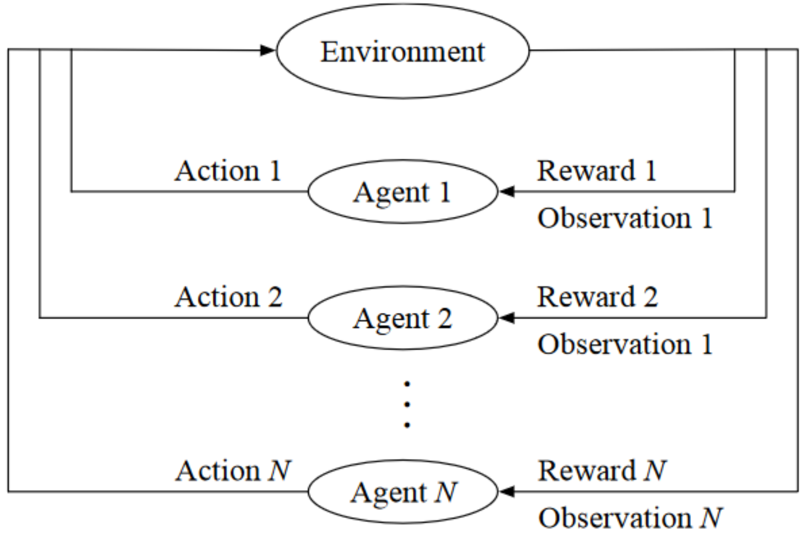
- Policy Gradient
- AC
- PPO

Multi-agent reinforcement learning

MARL tasks : the cooperative, competitive and mixed setting

In the cooperative setting, the agents work together to reach a common goal.

- combinatorial nature
- partial observability



POMDP

➢ Multi-agent reinforcement learning

MARL tasks : the cooperative, competitive and mixed setting

In the competitive setting, each agent has its own reward function and acts selfishly to maximize only its own expected cumulative reward.

zero-sum game

	A	*B*	*C*
1	30, -30	-10, 10	20, -20
2	10, -10	20, -20	-20, 20

~~Game Theory~~

Agent number
is small (often 2)

In the mixed setting, in the most general case each agent has an arbitrary but agent-unique reward function.

SEQUENTIAL DECISION-MAKING

➢ Multi-agent reinforcement learning

A SELECTED SUBSET OF RESEARCH AREAS AND RECENT ALGORITHMS OR MODELLING FRAMEWORKS FOR MULTI-AGENT SYSTEMS.

Category	Framework	Methodology
RL	Q-Learning [40]	Learns a tabular value function for optimal control with finite state and action spaces.
	Value function approximation [41], [42]	Scales to large / continuous state spaces by learning approximated value functions.
	Policy gradient methods [44], [55]	Scales to large / continuous state and action spaces by iteratively improving a policy.
MARL	Independent Learning [54]	Applies single-agent RL to each agent directly, ignoring the multi-agent aspect.
	Parameter Sharing [53]	Scales to arbitrary numbers of agents by learning a single shared policy for all agents.
	QMIX [60]	Decomposes and learns the joint-value as a monotonic combination of single-agent values.
	MADDPG [62]	Policy gradient method taking into account actions of other agents in a centralized critic.
	COMA [63]	Adds a counterfactual baseline for policy advantage estimation using a centralized critic.
	MAPPO [57]	PPO [55] via independent learning, achieves state-of-the-art results [20], [56]–[58].

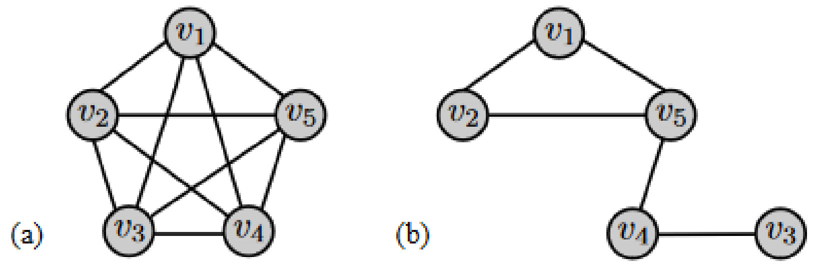
CONTENTS

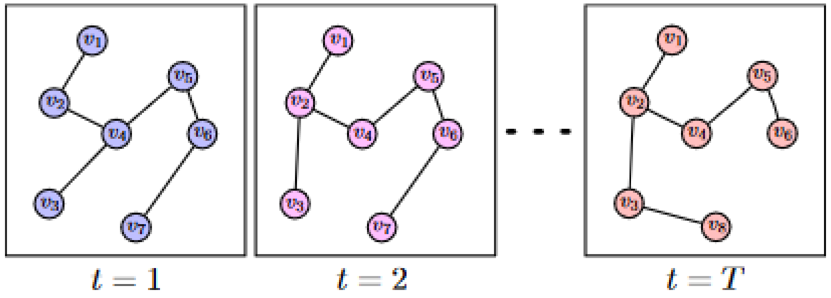
LARGE-POPULATION SYSTEMS

➢ Graph-based methods

- Factorized models
  a. Factored MDPs
  b. Partially-observed models
  c. Other scalable methods
- Complex network models

Visualization of (a) a fully connected graph of a system of 5 agents labeled as vi and (b) a coordination graph depicting local interactions in the system. The coordination graph allows factorization of the reward function into local factors and provides tractable solutions.





Visualization of an adaptive network over time. At time t = 2, the connection (edge) between nodes v3 and v4 ends. Until time t = T , node v7 leaves the network and node v8 makes a connection with node v3.

LARGE-POPULATION SYSTEMS

- Mean-field limits
  - Mean-field games
  - Mean-field control
  - Graphs and partial observability

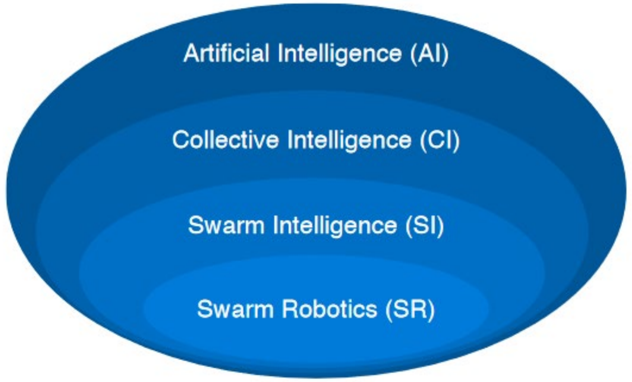
Pictorial scheme of approximation for mean-field games and meanfield control. The finite N -agent system is first approximated by a meanfield system, which is then solved through learning algorithms, thereby circumventing the difficult solving of the finite system. The resulting solution will be an approximately optimal solution in sufficiently large finite systems.

$$\begin{tikzcd}
\text{Cooperative Pareto optimum} & \text{MFC} \arrow[l, "\text{approx.}"'] \\
N\text{-agent system} \arrow[u, "\text{Optimize (coop.)}"'] \arrow[r, "N\to\infty"] \arrow[d, "\text{Optimize (comp.)}"] & \text{Mean-field system} \arrow[u, "\text{Optimize (coop.)}"'] \arrow[d, "\text{Optimize (comp.)}"] \\
\text{Competitive Nash equilibrium} & \text{MFG} \arrow[l, "\text{approx.}"']
\end{tikzcd}$$

LARGE-POPULATION SYSTEMS

➢ Collective swarm intelligence

- Reinforcement learning for swarm intelligence
- Swarm intelligence for decision-making



➢ Partial observability and decentralization

Partial observability and decentralization are a key element in large-population systems, both from the theoretic and applied point of view. Without partial observability, each agent must know the global state of the entire system and may thus coordinate perfectly through a global policy shared by all agents. Thus, there cannot be decentralization.

CONTENTS

APPLICATIONS

- Distributed computing
- Cyber-physical systems
- Autonomous mobility and traffic control
- Natural and social sciences

CONTENTS

FUTURE DIRECTIONS

- The limiting mean-field regime
- Higher-order complex networks
- Intersectional and application-oriented work

Mean-field game theory

Mean-field game theory is the study of strategic decision making by small interacting agents in very large populations.

Use of the term "mean field" is inspired by mean-field theory in physics, which considers the behaviour of systems of large numbers of particles where individual particles have negligible impact upon the system.

BACK

Nash equilibrium

In game theory, the Nash equilibrium is the most common way to define the solution of a non-cooperative game involving two or more players, each player is assumed to know the equilibrium strategies of the other players, and no one has anything to gain by changing only one's own strategy.

Standard prisoner's dilemma payoff matrix

B A	B stays silent	B betrays
A stays silent	-2 -2	0 -10
A betrays	-10 0	-5 -5

Pareto optimality

Pareto optimality is a situation where no individual or preference criterion can be made better off without making at least one individual or preference criterion worse off.

BACK