HUMAN-MACHINE INTERACTIONS IN MULTI-AGENT REINFORCEMENT LEARNING

University at Buffalo
The State University of New York

State University of New York at Buffalo
Alina Vereshchaka
March 21, 2022

ML

1

Human and machine teams in operations

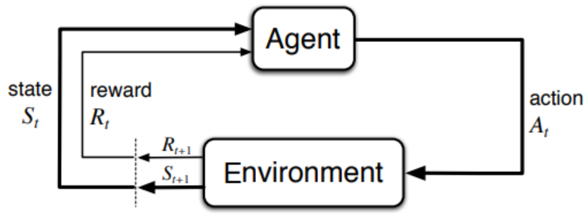
- RL methods have been successfully applied to various domains including robotics, transportation, autonomous driving, industry automation and epidemic mitigation.
- Many real-world problems require multiple RL agents to cooperate with each other. These problems fall into the realm of multi-agent reinforcement learning (MARL)
- Combining human and machine teams in operations can significantly increase efficiency
- However, there are a number of critical questions to consider

2

Markov Decision Process

Markov decision process (MDP) defined by the tuple $\langle s, a, P, O, r, \rho_0, \gamma \rangle$, where

- $s \in S$ denotes states, describing all possible configurations;
- $a \in A$ denotes actions;
- $P : S \times A \times S \to \mathbb{R}$ is the states transition probability distribution;
- O is a set of observations;
- $r : S \to \mathbb{R}$ is the reward function;
- $\rho_0 : S \to [0, 1]$ is the distribution of the initial state s_0;
- $\gamma \in [0, 1)$ is a discount factor



3

Stochastic Game

Definition

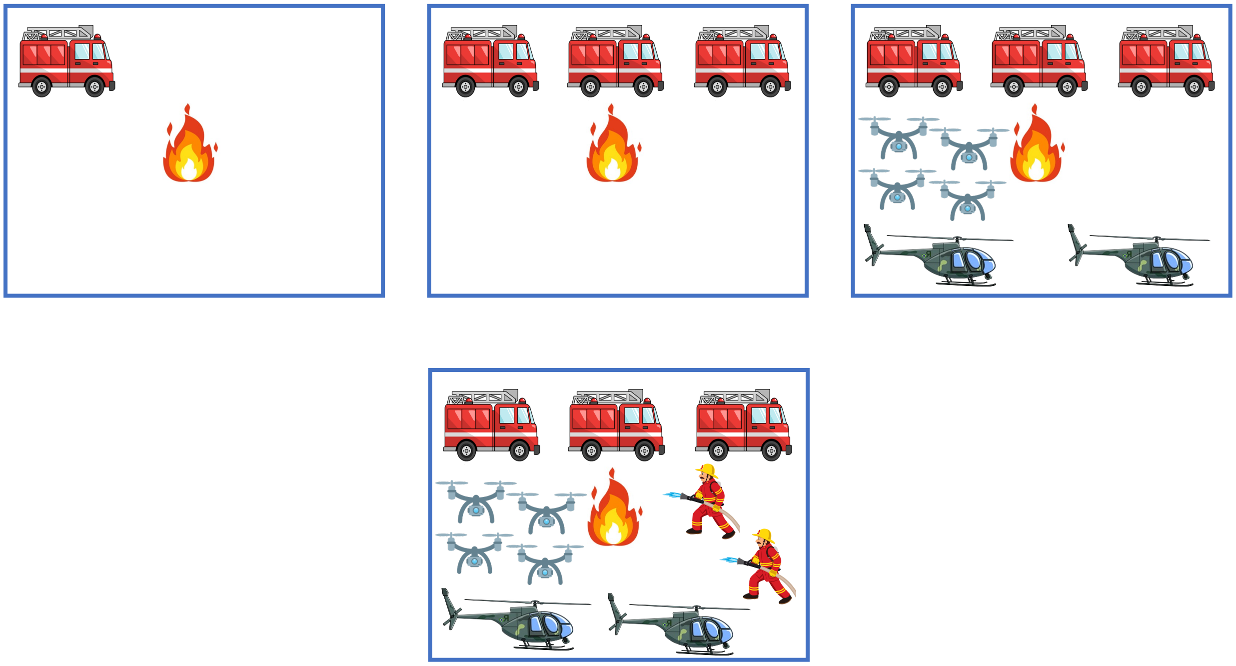
Normal-form game consists of:

- Finite set of agents $i \in \mathcal{N} = \{1, \cdots, n\}$
- Each agent $i \in N$ has a set of actions $A_i \in \{a_1, a_2, \cdots\}$
- Set of **joint** actions $A = a_1 \times a_2 \times \cdots \times a_n$
- Rewards function $r_i : A \to \mathbb{R}$, where $A = A_1 \times \cdots \times A_n$

Each agent i selects policy $\pi_i : A_i \to [0, 1]$, takes action $a_i \in A_i$ with probability $\pi_i(a_i)$, and receives reward $r_i(a_1, \cdots, a_n)$. Given policy profile $(\pi_1, \cdots, \pi_n)$, expected reward to i is

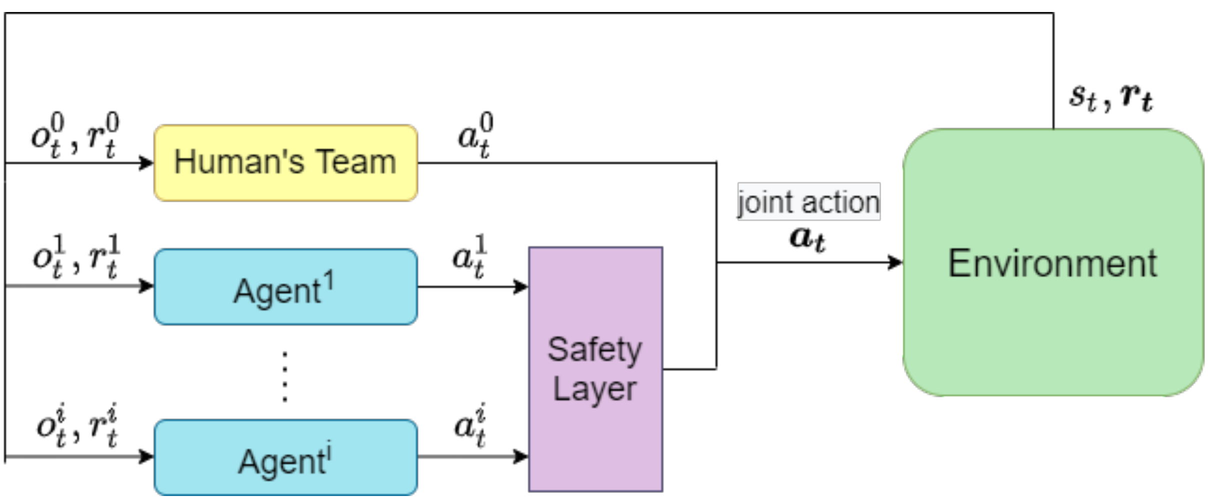
$$r(\pi_1, \cdots, \pi_n) = \sum_{a \in A} \pi_1(a_1) * \cdots \pi_n(a_n) * r_i(a)$$

4



5

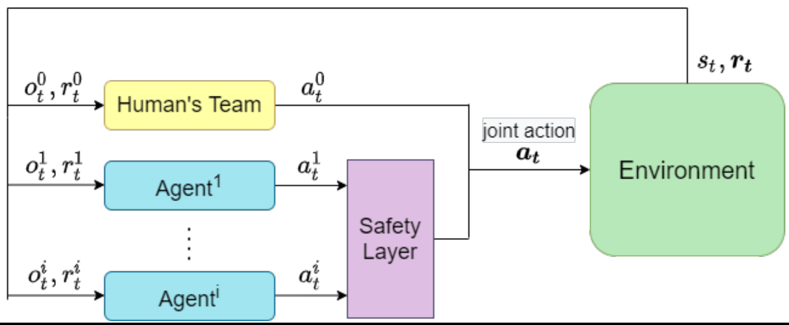
Multi-agent reinforcement learning framework with a human-agent collaboration



6

Setup Strategies

- Define the main objective for each type of agents, given their different observations and skills set.
- Explicitly provide their set of priorities and all of the necessary safety measurements while training the agents.



7

Critical questions

1. Ensure that each agent is trained to follow the safety constraints, including preventing the agent from damaging itself, preventing any damage to the environment in which they are navigating, and preventing any kind of harm to the human team.
2. Ensure that the combined teams of humans and agents are working in a collaborative environment, in which they are all aiming to solve a common goal, e.g. completing a rescue operation
3. In multi-agent reinforcement learning environments, we can assign different levels of priorities to our agents. In cases of human-machine collaboration, we aim to assign the main priority to the team of humans, and assign them with a higher priority in terms of action execution.
4. Agents optimize the objectives that are set by humans, so the agent may incorporate bias given the objective, the task formulation or the data used to train them.

8