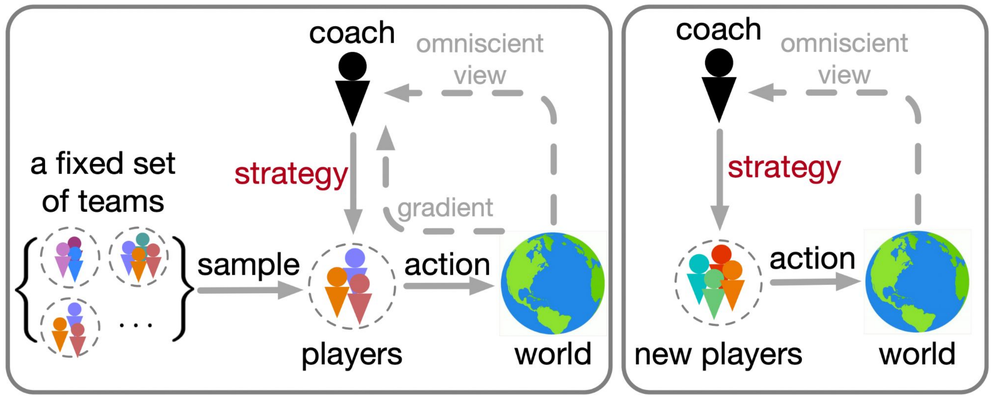
Coach-Player Multi-agent Reinforcement Learning for Dynamic Team Composition

Bo Liu[1], Qiang Liu[1], Peter Stone[1], Animesh Garg[2,3], Yuke Zhu[1,3], Animashree Anandkumar[3,4]
[1]The University of Texas at Austin, [2]University of Toronto, [3]Nvidia, [4] California Institute of Technology

Motivation

In practical multi-agent systems, agents with different characteristics may come and go. We investigate how to coordinate such teams effectively.

Coach and Player

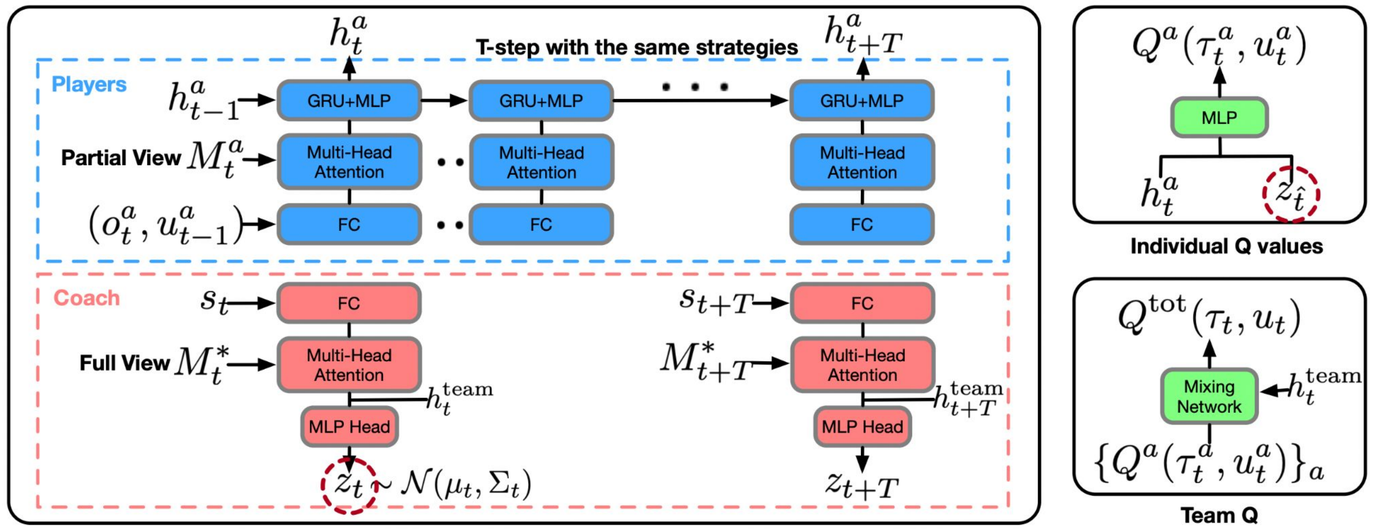


(a) Training (b) Zero-shot generalization

Coach (omniscient view)
- has omniscient view of the world
- broadcast *strategies* to agents once in a while

Players (partial views):
- have partial view of the world
- make decisions based on the most recent strategies

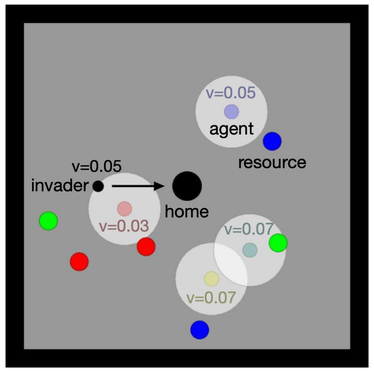
COPA model



Regularization: strategies should be identifiable from agents' behaviors.

Comm. frequency: broadcast only when the new strategies are different.

Results



Method	Env. ($n=5$)	Env. ($n=6$)	Env. (varying n)	f
Random Policy	6.9	10.4	2.3	N/A
Greedy Expert	115.3	142.4	71.6	N/A
REFIL	90.5±1.5	109.3±1.6	61.5±0.9	0
A-QMIX	96.9±2.1	115.1±2.1	66.2±1.6	0
A-QMIX (periodic)	93.1±20.4	104.2±22.6	68.9±12.6	0.25
A-QMIX (full)	157.4±8.5	179.6±9.8	114.3±6.2	1
COPA ($\beta=0$)	175.6±1.9	203.2±2.5	124.9±0.9	0.25
COPA ($\beta=2$)	174.4±1.7	200.3±1.6	122.8±1.5	0.18
COPA ($\beta=3$)	168.8±1.7	195.4±1.8	120.0±1.6	0.13
COPA ($\beta=5$)	149.3±1.4	174.7±1.7	104.7±1.6	0.08
COPA ($\beta=8$)	109.4±3.6	130.6±4.0	80.6±2.0	0.04

QR code (paper)

38th International Conference on Machine Learning, 2021