Deep Reinforcement Learning for Green Security Games with Real-Time Information

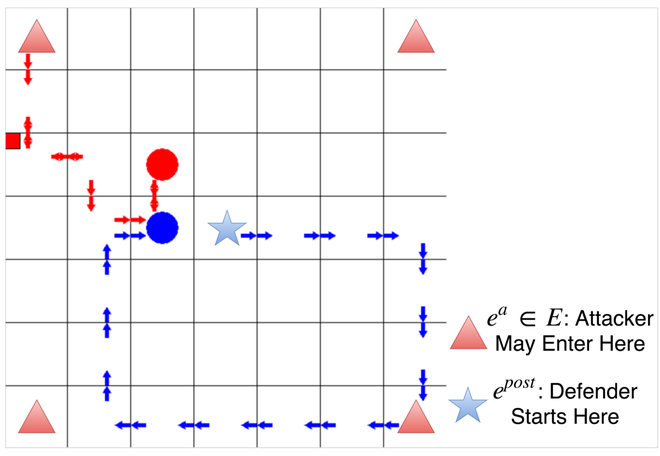
Yufei Wang[1], Zheyuan Ryan Shi[2], Lantao Yu[3], Yi Wu[4], Rohit Singh[5], Lucas Joppa[6], Fei Fang[2]
[1]Peking University, [2]Carnegie Mellon University, [3]Stanford University
[4]University of California, Berkeley, [5]World Wild Fund for Nature, [6]Microsoft Research

Introduction



Real-time information such as footprints and agents' subsequent actions upon receiving the information, e.g., rangers following the footprints to chase the poacher, are neglected in previous work of Green Security Games.
Our paper fills the gap. First, we propose a new game model GSG-I which augments GSGs with sequential movement and the vital element of real-time information.
Second, we design a novel deep reinforcement learning-based algorithm, DeDOL, to compute a patrolling strategy that adapts to the real-time information against a best-responding attacker.

GSG-I Game Model

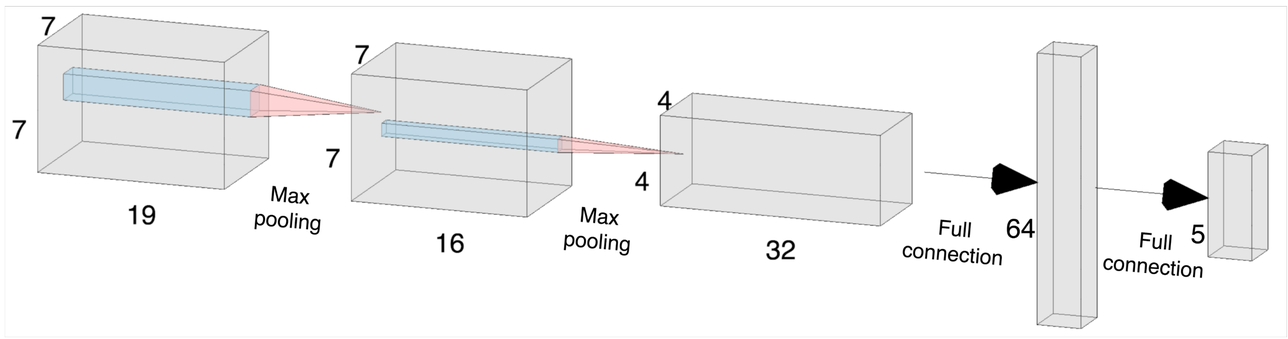


Attacker: red circle. He tries to put some attacking tools in the world to maximize the damage, meantime avoid defender.
Defender: blue circle. She tries to catch the attacker and remove the attacking tools as soon as possible.
Sequential Interaction: at each time step, attacker and defender both chooses a direction to move. Attacker also decides whether to put an attacking tool. The game environment decides whether an attacking tool will successfully launch an attack.
Footprints: red/blue arrows. Players leave footprints as they move.
Local observation: Players only observe opponent's footprints in the current cell.

DeDOL Overview

DeDOL builds upon the double oracle (DO) framework and the policy-space response oracle (PSRO). It starts with a restricted game and iteratively adds best response strategies to it, which is approximated by training DQN. Exploring the game structure, DeDOL uses domain-specific heuristic strategies as initial strategies in PSRO, and constructs several local modes for efficient and parallelized training.

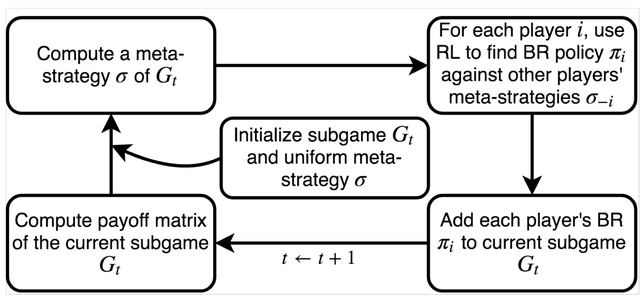
Approximating best response strategy against a fixed opponent by DQN



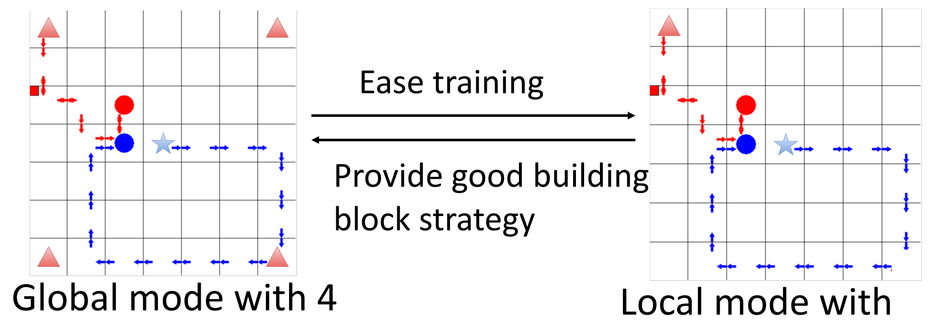
Input : a 3-D tensor with the same height and width as the grid world, with each channel encoding different information such as the observed footprints, and etc.
Output: each output unit represents the Q-value of choosing that action.

DeDOL: Computing Optimal Patrol Strategy by Enhancing PSRO

Vanilla PSRO



Local Mode



Global mode with 4 entry points

Local mode with top-left entry point

Doman-specific initial strategy

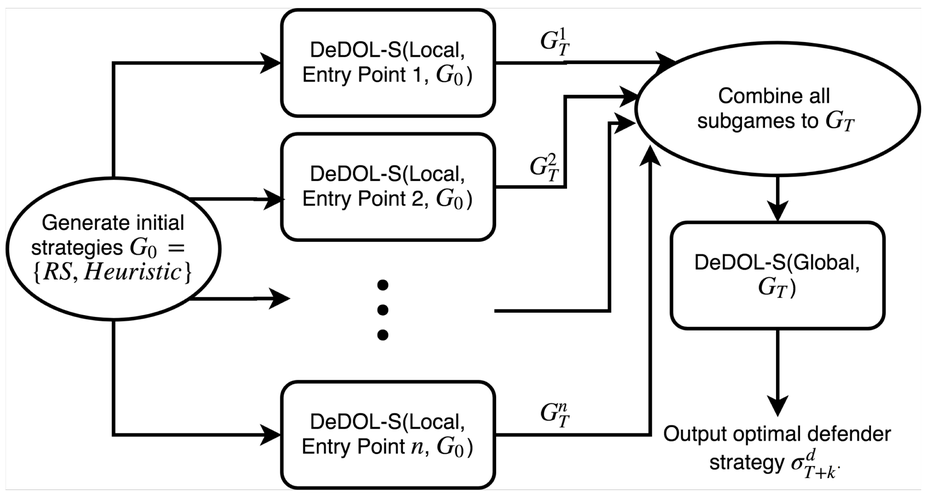
Attacker: parameterized heuristic

$$\pi_a(a_t^a = k|s_t^a) = \frac{\exp(w_p \cdot \bar{P}_k + w_i \cdot I_k + w_o \cdot O_k)}{\sum_z \exp(w_p \cdot \bar{P}_z + w_i \cdot I_z + w_o \cdot O_z)}$$

$$\eta_a(b_t^a = 1|s_t^a) = \frac{\exp(P_{m,n}/\tau)}{\sum_i \sum_j \exp(P_{i,j}/\tau)}$$

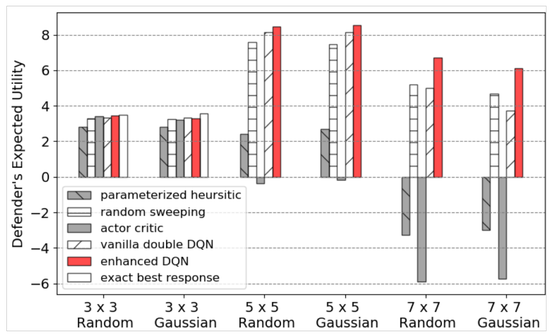
Defender: random sweeping. She moves along the boundary to find outgoing attacker footprints to follow. If multiple footprints, she randomly chooses one direction.

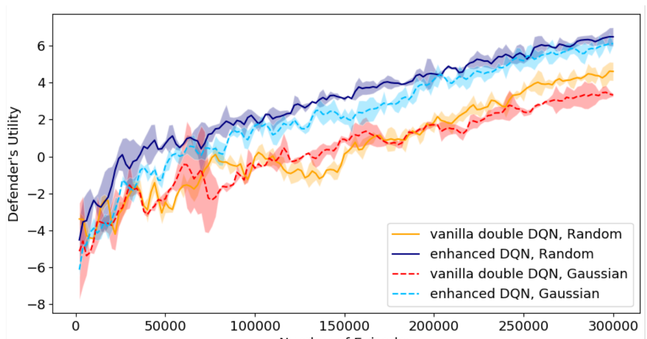
DeDOL workflow

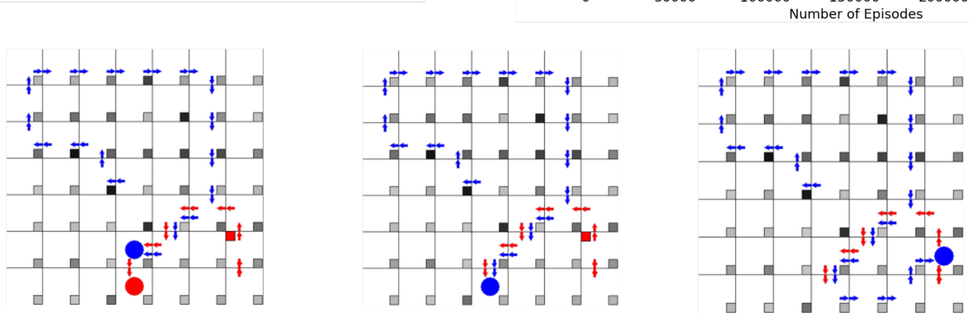


Experimental Results

Best Response Approximation







Expected Utility against a best-responding poacher

	Random Sweeping	Vanilla PSRO	DeDOL Pure Global Mode	DeDOL Local + Global Mode	DeDOL Pure Local Mode	CFR
3×3 Random	-0.04	0.65 (16)	0.73 (16)	**0.85** (10 + 2)	0.71 (20)	**1.01** (3500)
3×3 Gaussian	-0.09	0.52 (16)	0.75 (16)	**0.86** (10 + 2)	0.75(20)	**1.05** (3500)
5×5 Random	-1.91	-8.98 (4)	-1.63 (4)	-0.42 (4 + 1)	**-0.25** (5)	-
5×5 Gaussian	-1.16	-9.09 (4)	-0.43 (4)	**0.60** (4 + 1)	-2.41 (5)	-
7×7 Random	-4.06	-10.65 (4)	-2.00 (4)	**-0.54** (3 + 1)	-1.72(5)	-
7×7 Gaussian	-4.25	-10.08 (4)	-4.15 (4)	**-2.35** (3 + 1)	-2.62(5)	-

Acknowledgment

The Azure resources are provided by Microsoft for Research AI for Earth award program.