Carnegie Mellon University
The Robotics Institute

Akshat Agarwal,
Ryan Hope,Katia Sycara

Github Code
https://goo.gl/MUt6d7

Paper
https://goo.gl/k5mkdk

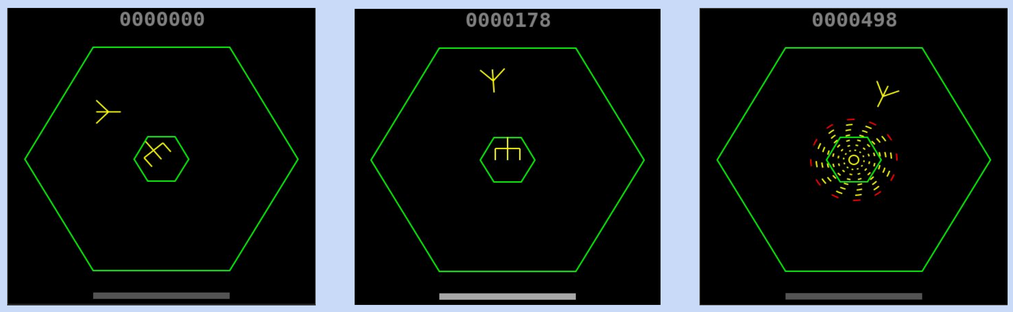
Challenges of Context and Time in Reinforcement Learning Introducing Space Fortress as a Benchmark

Why do we need context awareness?

- For any AI agent operating in the real world, optimal behavior can depend greatly on a broader context
- A shift in context can happen abruptly with little to no change in the observed state, but **require completely different strategies**
- Necessitates (1) learning to identify critical points where context changes, and (2) learning the different behaviors suitable for each context

- e.g. driving at 90 kmph usually safe, but would be very dangerous on a foggy day with icy road conditions (different context but difficult to observe in fog)
- Current RL benchmarks (Atari, Mujoco) have not specifically focused on context sensitivity
- Hence, we introduce Space Fortress as a new RL testbed for developing context aware RL algorithms

The Space Fortress (SF) environment

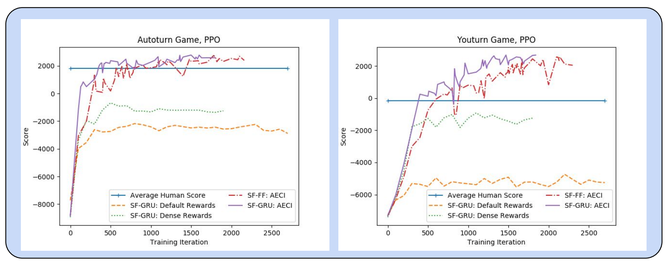


- Player/AI agent controls a ship which flies in a frictionless arena
- Fires missiles to destroy a fortress located centrally within the arena, incurring a penalty of -2 for each
- Hitting the walls or being hit by shells fired by fortress -> ship death, incurring a penalty of -100
- Destroying the fortress requires a context aware strategy, where the ship must fire its missiles slowly (<4Hz) till the fortress gets hit 10 times and becomes vulnerable, after which it can be destroyed *only by a rapid double shot* (>4Hz), giving a reward of +100
- The **fortress' vulnerability is the latent context** on which the agent's shooting strategy depends
- A simpler version of the game where ship orients itself automatically towards the fortress referred to as "**Autoturn**", while original game is "**Youturn**"

Experiments and Results

S.No.	Condition	Game Score	Fortress Deaths
1	Humans	216	14.3
2	RL: Default Rewards	-5269	0
3	RL: Dense Rewards	-1435	0.9
4	RL: AECI	**2356**	**41**

Table reports average result for (1) human evaluators, and (2-4) the best of {PPO, A2C and Rainbow} trained using 3 different sets of rewards, for 45M steps on Youturn



1. Human Evaluation: 117 people played 20 games of Space Fortress, given rules of the game beforehand
2. RL, Default Rewards: clipped rewards of -0.05 (missile penalty), -1 (ship death) and +1 (fortress death).
3. RL, Dense Rewards: introduce additional reward of +1 each time the fortress is hit by a missile.
4. RL, After Making Context Identification Easier (AECI): introduce additional reward of +/-1 for unit increase/decrease in fortress vulnerability, and a bonus of +2 (after clipping) for fortress destruction. These rewards help the agent to identify the critical points when the context (vulnerability) changes.

Conclusion

- No state of the art RL algorithm can learn to play Space Fortress, even with dense rewards.
- Making context identification easier through specific alterations of the reward structure allow PPO to achieve superhuman performance.
- Context insensitivity is the primary reason behind the inability of RL algorithms to learn to play SF
- Hence, Space Fortress (w/o modifications to its reward structure) is a challenging and interesting testbed for development of context-aware RL algorithms.

This research was sponsored in part by AFOSR grant FA9550-15-1-0442. The collection of human data and development of the OpenAI Gym interface for Space Fortress by Shawn Betts was supported by ONR grant N00014-15-1-2151.