Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play

Sainbayar Sukhbaatar¹, Zeming Lin², Ilya Kosrikov³, Gabriel Synnaeve⁴, Arthur Szlam⁵, Rob Fergus¹²
¹New York University ²Facebook AI Research

Motivation
Training an agent requires a lot of reward signals. But often external rewards are expensive to obtain. Can an agent learn about its environment without external rewards? Let the agent generate its own task and rewards!

Approach
• Let the agent play an imitation game with itself
• Single agent, but two separate minds

Two versions of the game
First, Alice acts freely in the environment until she stops. Then, Bob takes control and asked to reach the initial state where Alice started.

Internal reward during self-play
Alice wants to find simplest tasks where Bob struggles. Easier for Bob to learn since it’s only just beyond his current capabilities. Automatic curriculum.

Experiments (code available: http://cims.nyu.edu/~sainbar/selfplay)
Toy example: Long hallway
• Learn to navigate in a long corridor
• Reverse self-play
• Simple tabular policies

MazeBase: LightKey task
• Small 2D grid separated into two rooms by a wall
• The grid is procedurally generated (object/agent locations are randomized for each episode)

Target task is to reach the goal flag in the opposite room when light is off and door is locked.

Conclusion & Future directions
• An intrinsic motivation method for learning transitions between states
• Works with discrete and continuous environments
• A novel way to use self-play in a single agent environments
• In future: self-play in abstract state space, option discovery, different game

Related Work
• Robust Adversarial RL (Pinto et al., 2017)
• Automatic Goal Generation (Held et al., 2017)
• Hindsight Experience Replay (Andrychowicz et al., 2017)
• Reverse Curriculum Generation (Florensa et al., 2017)

Baseline: VIME (Houthooft et al., 2016), SimHash (Tang et al., 2017)

Toy example: Long hallway

Baselines: count-based exploration \( \alpha / \sqrt{N_{t}} \)

Target task is to reach the same state as Alice (position + velocity)

Box2D: MountainCar
• The target task is to reach mountain top
• Bob’s goal in self-play is to reach the same state as Alice (position + velocity)

Mujoco: SwimmerGather
• The target task is to swim and eat green apples
• Bob’s goal in self-play is to return to Alice’s initial position (reverse self-play)

StarCraft: Build Marines
• Control multiple units
• The target task is to build marines in given time
• Bob’s goal in self-play is to build as much stuff as Alice (ignore positions)