Deep Reinforcement Learning with POMDPs
Maxim Egorov

Learning Framework

Memory
Manages the experience dataset of the agent. Supports sampling and quick memory access

Simulator
Simulates the environment and generates experience tuples \((s,a,r,s')\). Currently handles Atari, Agar, and POMDPs.

Learner
The brain of the framework. Approximates Q-values during training, and outputs optimal actions during evaluation

POMDPs
Partially observable MDPs (POMDPs) are a framework for decision making under uncertainty. The state in POMDPs is partially observable. Agents can make decisions based on their action-observation histories or on their beliefs.

Mixed observability MDPs have fully observable variables in the state.

Deep Learning
Use a Deep Q-Network (DQN) to approximate Q values for a given belief or action-observation history

Results

Tiger
Objective: Escape and avoid the tiger

Rock Sample
Objective: Sample rocks that have scientific value

Conclusion
The contributions of this work are:
1. Created framework for solving POMDPs with DQNs
2. Extended approach to problems with mixed observability
3. Used SOM with convolutional neural networks to improve convergence rates

Future work will investigate how belief structure and convergence rates are related. Code will be open-sourced soon.