Policy Consolidation for Continual Reinforcement Learning

Christos Kaplanis[1], Murray Shanahan[1,2] and Claudia Clopath[1]

[1]Imperial College London, [2]DeepMind

11th June 2019

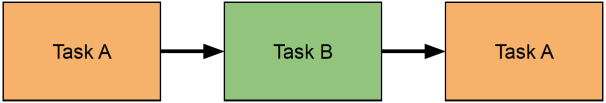
Motivation

Motivation

- Catastrophic Forgetting in Artificial Neural Networks

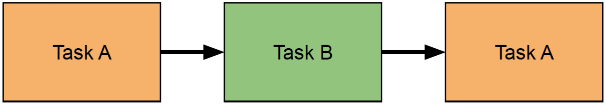
Motivation

- Catastrophic Forgetting in Artificial Neural Networks



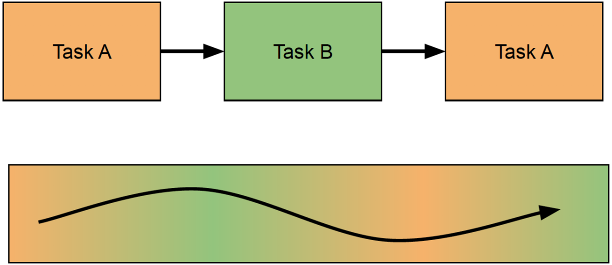
Motivation

- Catastrophic Forgetting in Artificial Neural Networks



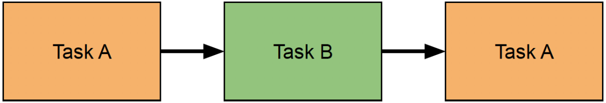
Motivation

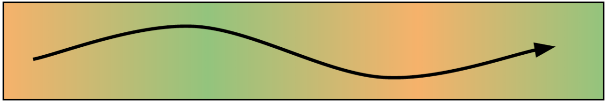
- Catastrophic Forgetting in Artificial Neural Networks



Motivation

- Catastrophic Forgetting in Artificial Neural Networks

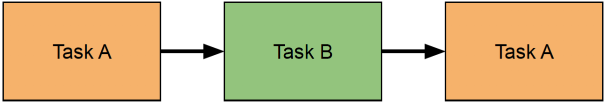


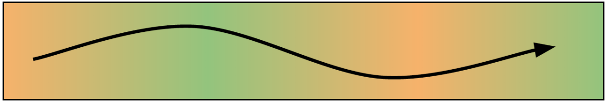


- Agents should cope with

Motivation

- Catastrophic Forgetting in Artificial Neural Networks

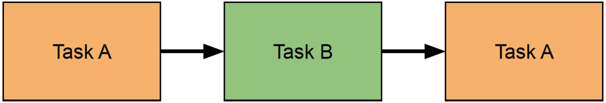


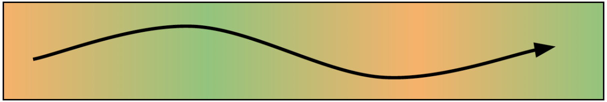


- Agents should cope with
  - Both discrete and continuous changes to data distribution

Motivation

- Catastrophic Forgetting in Artificial Neural Networks

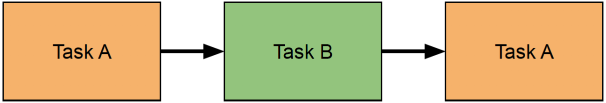


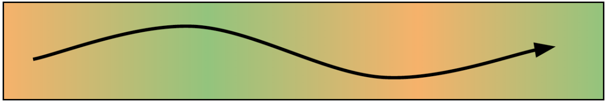


- Agents should cope with
  - Both discrete and continuous changes to data distribution
  - No prior knowledge of when/how changes occur

Motivation

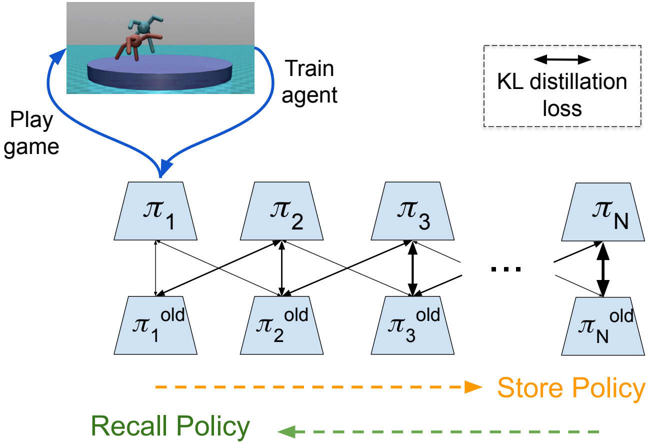
- Catastrophic Forgetting in Artificial Neural Networks



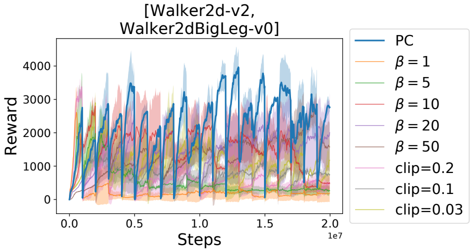


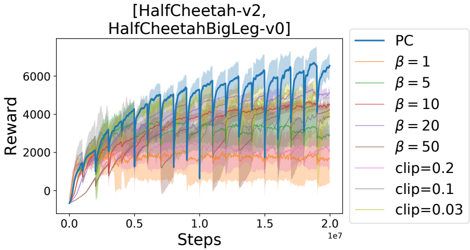
- Agents should cope with
  - Both discrete and continuous changes to data distribution
  - No prior knowledge of when/how changes occur
- Test beds: alternating task, single task and multi-agent RL

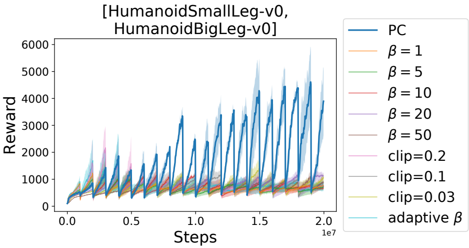
Policy Consolidation



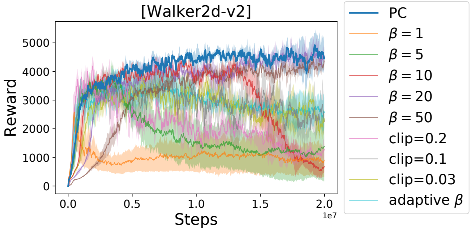
Alternating task experiments

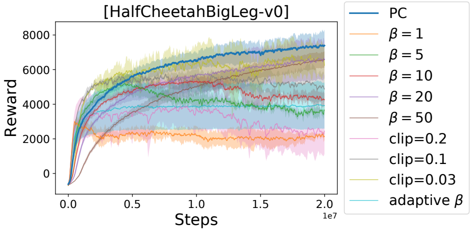


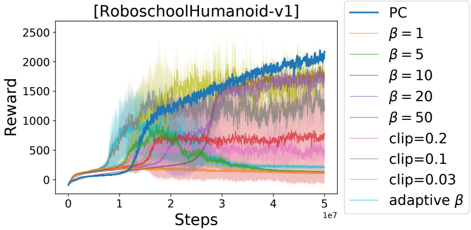




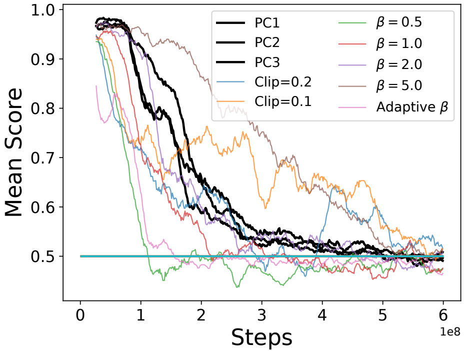
Single task experiments



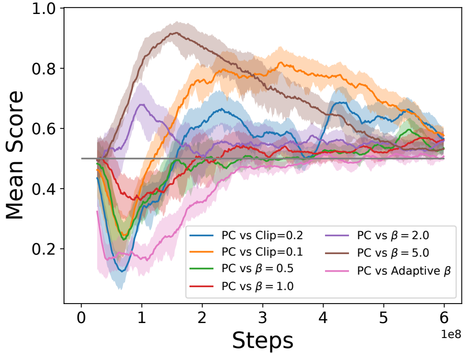




Multi-agent self-play experiments



(a) Final model vs. self history



(b) PC vs. baselines over training

Future work

Future work

- Prioritised consolidation

Future work

- Prioritised consolidation
- Adapt for off-policy learning