

## 進捗レポート (18T0006, 小室 光広)

### 今週の進捗サマリ

- シングルエージェントでの extended 実験
- マルチエージェント化の実装

## 1 進捗詳細

### 1.1 シングルエージェントでの extended 実験

2019 年 9 月 10 日に行われたゼミの際にご指摘いただいた結果が悪かったモデルについて追加で 400episode 学習を進める追加実験を行った。結果, pong の点数としては 21-6 でモデルの負けであるが, 1 点追加で点数を上げることができた。本追加実験では Random 値を固定しておらず, 比較のため Random 値を固定したシングルエージェントの実験を行う必要があるが, まだ行っていない。(実行自体はすぐできるため明日行う予定である) また, Seed を固定した場合における結果についても改めて考察することとする。

### 1.2 マルチエージェント化の実装

シングルエージェントで使用した Pong 用のサンプルコードを SeedFixed なマルチエージェントコードに書き換える作業を行っている。当初の案ではマルチエージェント化手法としてパイプライン的に行う *Sync*(仮称) とマルチスレッド的に行う *Async*(仮称) を想定していた。ここから大きく変わることは無いが, Sync モデルをメインに実装しており, 実装上 Sync モ

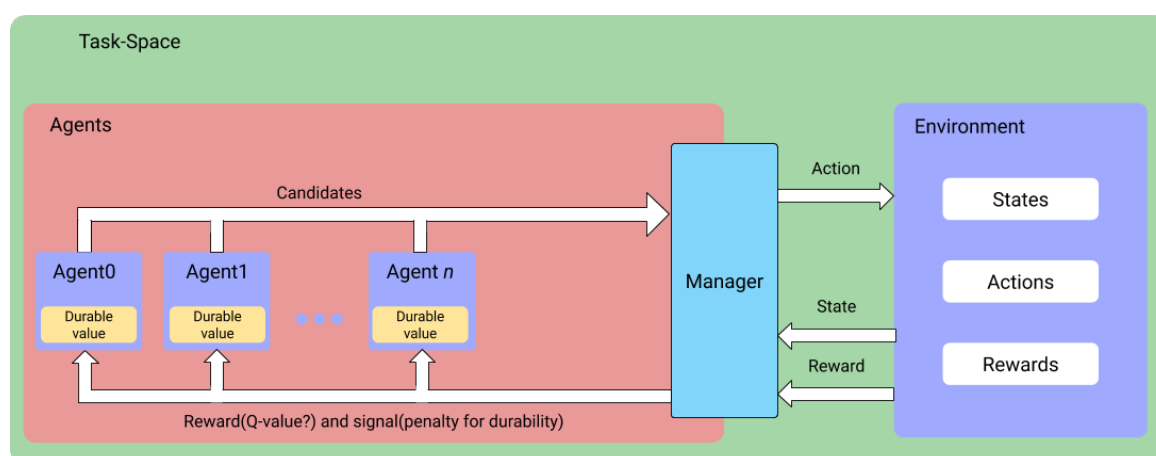


図 1: Sync(仮称) モデル

デルにおける Manager 及び Environment の関係について変更する必要が出てきた。変更点を以下に示す。

- 各 Agent0 から Agent n までそれぞれに Environment が必要
- Manager としていたものを `core_agent` として, 図 1 の Environment を `core_environment` とする

各 Agent から Action を候補として出す際, どうしても Environment の Step が進んでしまうため, 正しい学習とは言えなくなってしまう可能性がある, つまり Agent の Action n 個分だけ変化した Environment を Manager が扱う事になってしまう。この場合, FrameSkip に類似する現象と考えられるため, 精度低下が考えられる。次に Manager を Agent 化することについて, Manager の実装を考えたとき, 各 Agent から Best な Action を起こす Agent を選び, その Action を実際に Manager から Environment にやってほしいので, 実装上 Manager を `core_agent` として扱うこととした。

これらを踏まえ, 現在 Agent class 及び Environment class, Train step の関数までは実装できているが, 学習ループ中に RuntimeError で動作が停止することを確認しており, 修正中である。best Agent を耐久値 0 以下の Agent に Deep コピー

することで Agents の淘汰を行っているが、それが一因となっている可能性も考えられ、どこを修正すればいいのか手探りな状態であるため、時間を要すると考える。早期の解決を目指したい。

## 2 今後の予定

前述の通り、大まかな実装が完了したものの、Agent の淘汰等細かい部分でバグが発生しているため、実験まで到達していない状況である。9 月は残り 1 週となり非常に厳しい状況だが、今月中の実装を目指したい。また、次週ゼミまでに SeedFixed なシングル Agent のサンプルコードでの実験を行いたい。