協調型マルチエージェント強化学習による避難学習

1. 研究動機

SSI では学校から 1 人のみで避難することを学習させたが、現実の避難の際は大勢の人が同時に避難することになる。そこで、SSII では複数人の同時避難を学習させることにした。

2. 研究目的

・複数人が同時に行動する場合においても適切な避難方法を学習することは可能なのかを検証する。
・ML-Agents に新たに実装された協調型マルチエージェント強化学習トレーナー「MA-POCA」と従来のトレーナー「PPO」との学習状況の比較をする。

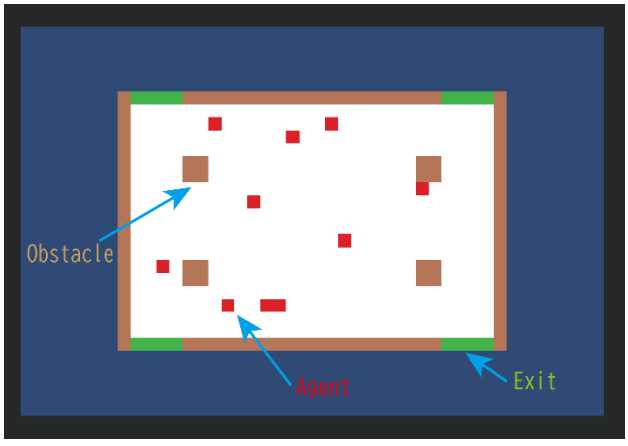
3. 協調型マルチエージェント強化学習とは？

従来の強化学習は、1 体の Agent（AI)が置かれた環境について観察・行動することで報酬を受け取り、その報酬を最大化するように試行錯誤しながら学習するものだった。これに対し協調型マルチエージェント学習では、いくつかの Agent をまとめて 1 つのグループとして捉え、そのグループ全体に報酬が与えられる。

これにより、グループ全体として報酬をより多くもらえるように各々が協調的に試行錯誤して学習するようになる。

4. 研究方法

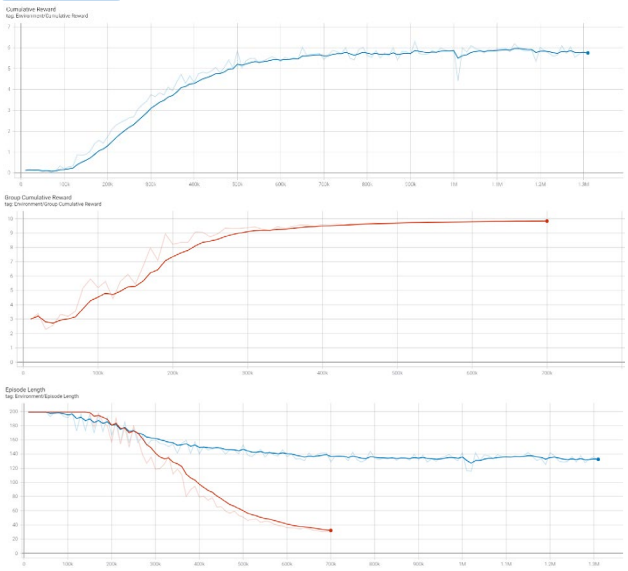
ゲームエンジン Unity とその強化学習ライブラリ ML-Agents を利用する。作成した学習環境の仕様は以下の通り。



【観察】周囲 11×11 マスに何があるかの情報を得る。
【行動】1 回の行動で前後左右に動く（またはその場にとどまる）ことが可能。
【報酬】Agent が 1 体 Exit に到達するごとに *+1.0*。
Agent が 1 回行動するごとに *-0.001*。
【終了条件】10 体全ての Agent が Exit に到達、又は各 Agent が 1000 回行動したとき。

この環境下で、従来のトレーナー「PPO」と協調型のトレーナー「MA-POCA」の 2 つでそれぞれ学習させる。

5. 結果



▲上：報酬推移グラフ（PPO） 中：報酬推移グラフ（MA-POCA） 下：避難完了までの時間推移グラフ

それぞれのトレーナーで学習が収束するまで学習をしたところ、PPO は 130 万エピソード（約 50 分）、MA-POCA では 70 万エピソード（約 2 時間）かかった。なお、どちらの場合も最終的に全 Agent の避難は達成できていた。※Agent の行動の様子は、右上の QR コードから動画で見ることができます。

6. 考察

MA-POCA の方が PPO よりも短い時間で避難完了できていることから、今回の避難学習では MA-POCA が適しているようだ。また、グループ内でフィールドの情報を共有しているので、Exit の位置がより迅速に把握できた結果、避難時間の短縮につながったのだと考えている。

7. 結論と今後の展望

今回の避難学習のように、多数の Agent が同じ目標をもって行動する場合、MA-POCA を用いることが効果的であるということが分かった。

今後は SSI の時と同様に、戸山高校の環境を用いて複数人による避難学習をしてみようと考えている。

8. 参考文献

・Unity ML-Agents Toolkit - Release 19
https://github.com/Unity-Technologies/ml-agents/tree/release_19
・ML Agents | 2.2.1-exp.1
https://docs.unity3d.com/Packages/com.unity.ml-agents@2.2/api/index.html