

学校内の避難における強化学習モデルの利用

東京都立戸山高等学校 SS 情報 1 年 柳瀬雄太

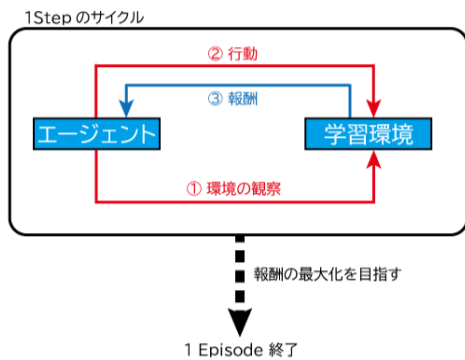
1. 研究動機

春休みに、YouTube で強化学習を用いた様々な検証動画を視聴し、強化学習というものに興味を持った。また、近年は災害が激甚化し、毎年のように日本のどこかで大きな被害が発生している。そこで、「人工知能は避難行動を学習し最適化できるのか」について疑問を持ち、研究してみようと考えた。

2. 研究目的

1. 強化学習によって人工知能に避難行動を学習させることはできるのか。
2. 1 が達成できた場合、人間側が想定するような振る舞いを見せるのか。

3. 強化学習とは？



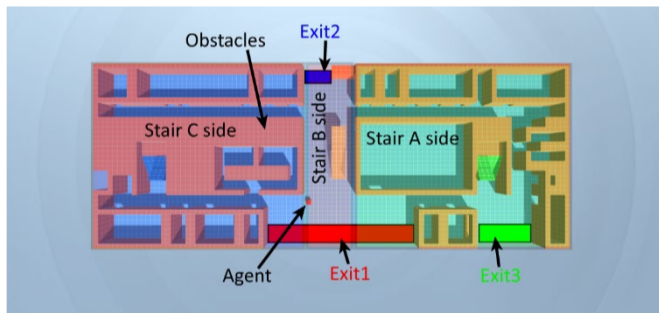
強化学習は人工知能(以下、Agent とする)に学習をさせる「機械学習」の手法の一つである。一般的な強化学習の流れを左図に示す。

Agent は 1 ステップごとに行動と学習環境の観察を行い、その行動に応じて報酬(正、負いずれかの数値)が与えられる。そして、報酬を最大化できるように行動を

最適化していき、事前に設定したステップが経過したら 1 回の学習を完了する(1 エピソード終了)。また、Agent がとる様々な行動に対してそれぞれどのくらいの報酬を与えるか、学習環境のどんな情報を観察させるかなどは人間側がプログラムを書いて指定する。

4. 研究方法

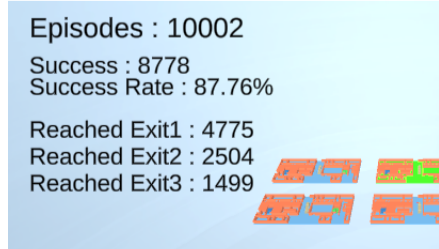
ゲームエンジン Unity と、その強化学習ライブラリ「ML-Agents」を使用して強化学習を行った。学習環境には私の通う戸山高校をモデルとした学校を使ったが、簡単のため 1 階部分のみを使用した。作成した学習環境の仕様を以下に示す。



学習環境には私の通う戸山高校をモデルとした学校を使ったが、簡単のため 1 階部分のみを使用した。作成した学習環境の仕様を以下に示す。

オブジェクト名	仕様	Agent への報酬
Agent	ランダムな位置に生成される。前後に移動、左右に回転する。自身の周囲 32 方向にレーザーを発射し、周囲環境を観察する。また、今の自分の座標と速度、角速度、Exit1~3 の座標(各 3 次元)を取得する。	毎ステップごとに -0.0002 の負の報酬
Exit	Agent が触れると、 避難成功 となる。	Exit に Agent が触れた場合、 +0.75 の正の報酬 を与え、エピソード終了。Agent の生成位置に最も近かった場合、さらに +0.75 を与える。
Obstacles	Agent が触れると、 避難失敗 となる。	Agent が触れた場合、 -1.0 の負の報酬 を与え、エピソード終了。

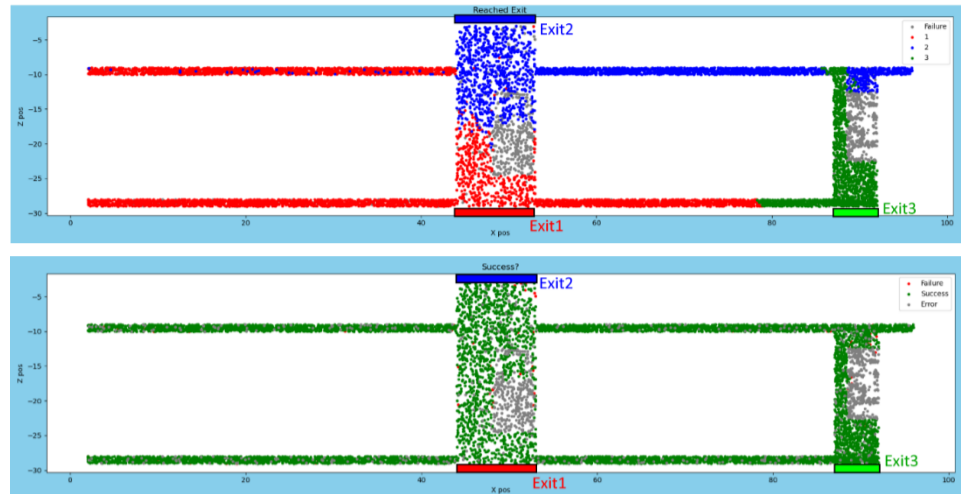
5. 結果



約 40 時間、1000 万エピソード学習した後、10000 回推論をさせた(右図)。ここでいう推論というのは、学習済みのモデルを使って、1 階のランダムな範囲から避難をさせてみることである。すると、9 割弱の確率で避難に成功していることが分かった。

6. 考察

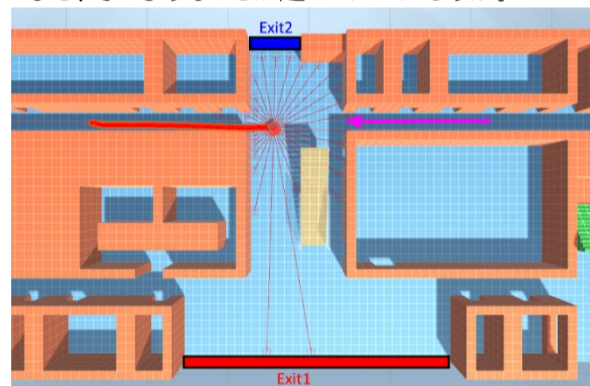
5 で分析した結果だけではわかりにくかったので、Python を用いてスタート位置の座標(x, z 座標)と避難成功判定を図にしたのが下図上段、到達出口を図にしたのが下図下段である。



成功を示す緑色が多くを占めているので、学習環境のほぼすべての場所から避難することができているといえる。また、壁沿いの地点で失敗を示す赤色が見られるが、これはスタート直後に観察のため回転をしたときに、壁に当たってしまったことが原因と考えられるため、無視してよい結果であろう。

また、到達出口に関しては、スタート地点から最も近い出口に避難していることがわかる。しかし、四角で囲まれた範囲では、仮説通りには行動しなかった。特に左上の部分の多くは、一番近いはずの Exit2 ではなく、Exit1 に避難している。

なぜ、そのようなことが起こったのだろうか。



推論時の Agent の動きから私は、Stair A side では、上部と下部で最も近い出口が異なるため、同じモデルの中で、2 方向の出口が最適化されて Agent が混乱した可能性を考えた。Stair A side 上部から Exit2 へ行くうえで、途中で Exit1 と Exit2 の両方にレーザーが当たってしまう。

学習初期の試行錯誤の段階で Exit1 に行った方が報酬をより多くもらえると判断してしまったのかもしれない。

一方で、Stair C side 上部から Exit2 へ行く上図のピンクのルートでは、レーザーは直前まで Exit2 にしか当たらないため、混乱は起きない。

7. 結論

まず、目的 2-1 に関しては、私のような個人研究レベルであっても、ライブラリを活用することで強化学習によって避難行動を学習させることは可能であることが分かった。また、目的 2-2 に関しては、おおむねこちらの予想通りの振る舞いを見せたが、予想外の行動を取る場面もあった。客観的に信頼できるようなモデルの作成にはまだまだ自身の技術が足りないことを感じた。

8. 今後の展望

残念ながら、機械学習のモデルは行動の理由を説明できないので、示された結果にどう信頼性を持たせるかが課題である。現状では、推論時のエージェントの動きの解析を詳細に行うしかないだろう。

今後は校舎 2 階以上からの避難の学習も考えている。3 か所にある階段をエージェントがどのように使うのか、とても興味がある。また、今回のようなシングルエージェントではなく、マルチエージェントによる避難学習も視野に入れている。

9. 参考文献

・「ML-Agents」の公式ドキュメント

https://github.com/Unity-Technologies/ml-agents/tree/release_17

・強化学習とは | データアーティスト株式会社 | AI(人工知能)

<https://www.data-artist.com/contents/reinforcement-learning.html>

・Unity ML-Agents 実践ゲームプログラミング v1.1 対応版(Unity ではじめる機械学習・強化学習)

著者:布留川 英一 出版社:ポーンデジタル