深層強化学習におけるライントレース制御に関する研究

足利大学　平石研究室

S18129　武田翔太

はじめに

- 本研究では、レゴマインドストームによる深層学習教材を利用して，ライントレースにおける深層強化学習の適用について検証を行なった。
- 教材は株式会社アフレル制作の「ロボットではじめる深層学習」を用いた。
- Pythonでシミュレータを動作させ、JupyterLab上で

ライントレース制御を学習させ、規定のコースを周回する精度を調べる。

卒業研究Aの内容

- シミュレータを使用し、学習モデルの訓練とその性能のテストを行った。
- 訓練を行う回数を少しずつ増やしていき、

回数ごとの結果を計測する。

- 卒業研究Aでは訓練回数を100回から1500回

まで100回刻みに行ったが、テスト用コースを

周回し続けることはできなかった。

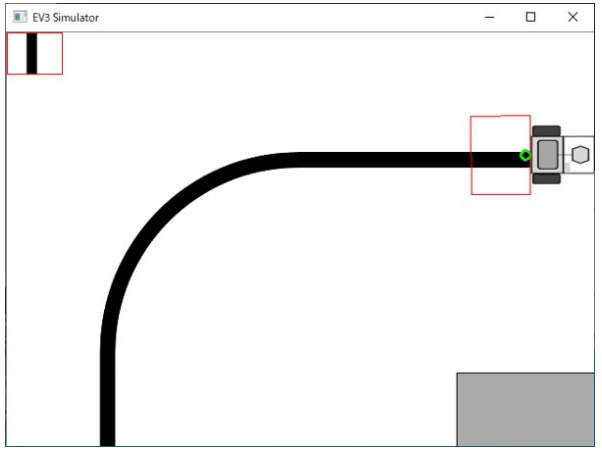

図　使用したシミュレータ
（訓練用コース）

目的

- 卒業研究Aではコースを周回し続けることができなかったため、コースを周回し続けられる訓練方法を探す。

シミュレータでは1000ステップ分ロボットが周回し続けると自動終了するように設定しているため、右図のコースを1000ステップ周回し続けることを目標とする。

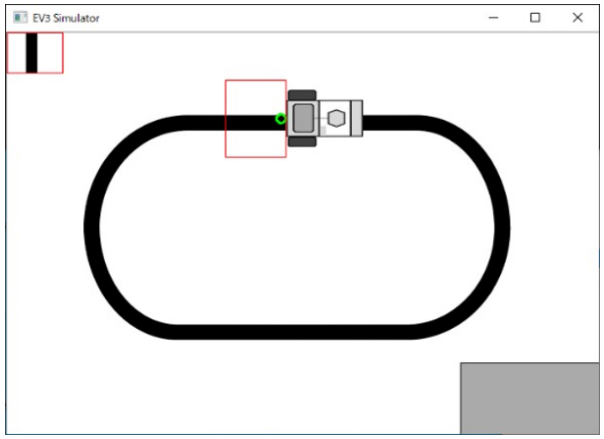

図　テスト用コース

実験方法

- 実験はJupter Lab上でプログラムを実行し、シミュレータ上のロボットを動かして行う。
- シミュレータ上の訓練用コースを規定ステップまでロボットを動かし続け、学習モデルの訓練を行う。
- 学習に使われる報酬はロボットの視界の中心から視界内の黒線の重心までの距離で決定している。
- 訓練された学習モデルを用いてテスト用コースをロボットに走らせ、脱落するまでのステップ数を計測する。

実験方法

- コースを周回し続けられる方法を調べるにあたり、下記の3つの訓練方法を実験し、どのような効果が得られるのかを調べた。

1. 卒業研究Aと同条件で訓練回数を増加させる
2. テスト用コースと訓練用コースを同一のものにして訓練する
3. 2.に加え、難易度が高いと考えられるカーブを重点的に訓練させる。

また、データの計測回数も10回から20回に増加させる。

実験1　同条件で回数を増加

- この実験では卒業研究Aの実験と同条件で訓練回数のみを増加させて実験を行う。
- 卒業研究Aでは100回から1500回まで100回ずつ増加させていた訓練回数を、1000回から5000回まで1000回ずつ増加させてデータを計測する。(1000回,2000回,3000回…となる)

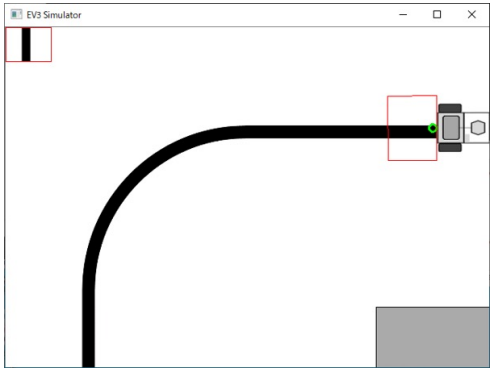

図　訓練用コース

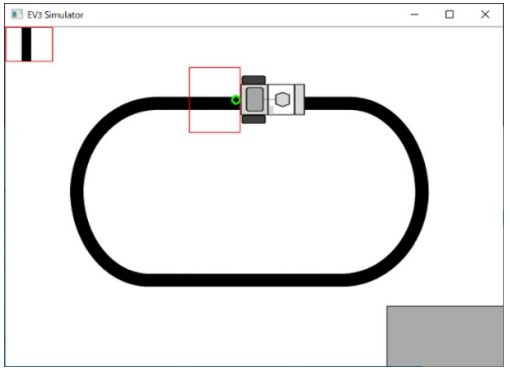

図　テスト用コース

実験２　テスト用と訓練用コースの統一

- 訓練用コースと実際にテストを行うコースを分けていたが、この実験では訓練をテスト用コースで行う。
- 条件をテストの際と同一にすることで、
  より効率的な学習が行えるかを調べる。

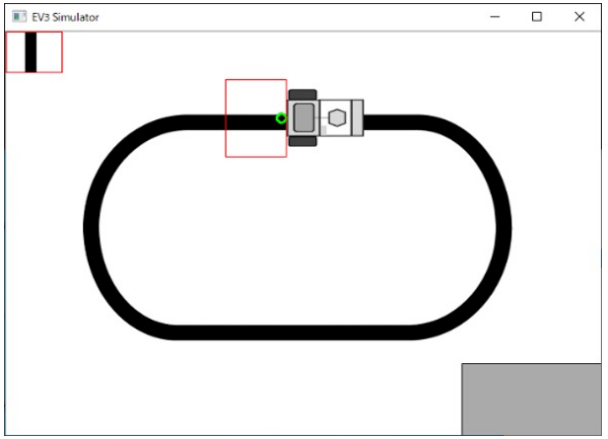

図　訓練・テスト用コース

実験3 訓練用コースの開始場所の変更

- コースの設定を変更し、カーブの途中からロボットをスタートさせるようにした。
- この変更により、必ずカーブから訓練が始まるため、カーブを重点的に訓練できるようになると考えられる。

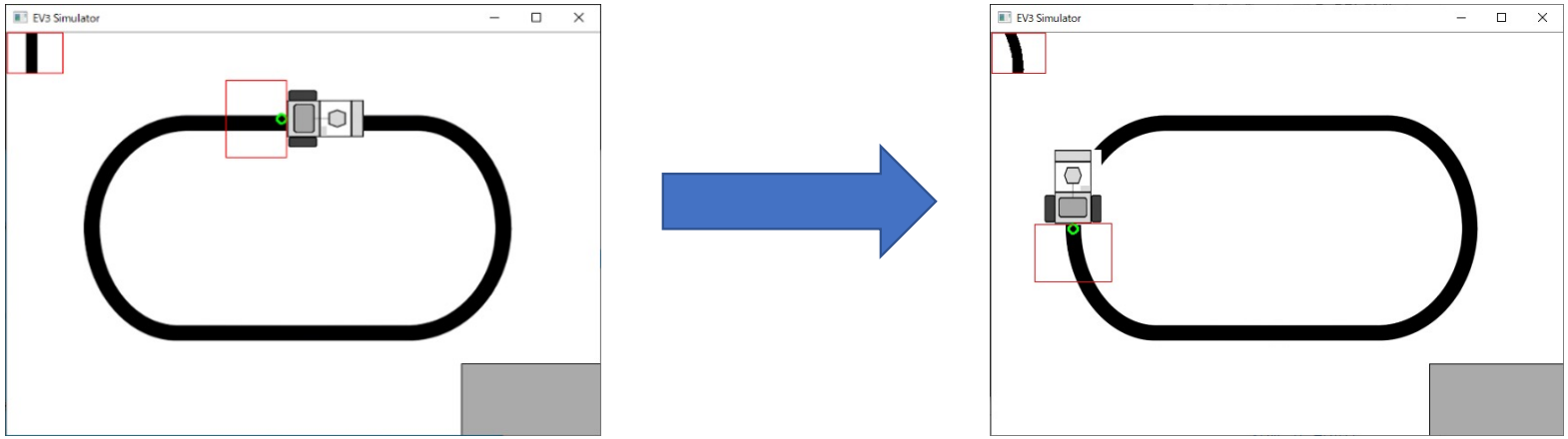

図　コースの開始地点を変更した図

実験１～３の結果

- 表はそれぞれの実験の結果をまとめたものである。
- 数値はロボットがコースを外れるまでのステップ数であり、20回の施行の平均を示している。
- 実験1~3ではコースを1000ステップ周回し続けられるものはなかった。
- 訓練回数を増加させることで数値が低くなる傾向があった。

表　実験1~3の平均ステップ数

訓練回数	平均ステップ数		
	実験1	実験2	実験3
1000	212.6	52.3	144.7
2000	118.2	40.8	47.0
3000	494.8	90.3	36.6
4000	42.0	74.1	56.0
5000	88.3	47.9	54.6

実験４　報酬量を変化させる

- 本研究ではロボットの視界の中心から視界内の黒線の重心までの距離Xで報酬を決定している。
- ラインから離れた場所（マイナスの報酬を受け取るが実験が終了しない地点）でライントレースを維持するより、素早くコースアウトした方が合計の報酬量が大きくなってしまう現象が起きていた。
- 報酬量を右表のように変更し、実験を行った。
- コースは実験2のものを用いた。

表　距離Xに対応する報酬量

距離X	報酬量	
	変更前	変更後
0≦X≦5	5	5
5＜X≦10	3	3
10＜X≦15	1	1
15＜X≦30	-2	0
30＜X＜40	-3	0
40≦X	-6	-6

実験4の結果

- これまで同様に20回の実験を行ったが、ここでは前半10回と後半10回に分けて右表に結果をまとめた。

- この結果を見ると、訓練回数が増加するにつれ、平均ステップ数も増加している。
- 特に前半の訓練モデルでは、5000回訓練した際には常にコースを周回し続けられるようになった。

表　実験4の平均ステップ数

訓練回数	平均ステップ数		
	前半	後半	平均
1000	44.3	103.1	73.7
2000	57.5	163.7	110.6
3000	316.3	219.9	268.1
4000	637	228.3	432.65
5000	1001	629	815

まとめ

- 今回の実験では、モデルを訓練させるコースを変更することではコースを周回させ続けることはできなかった。
- しかし、報酬量を変化させた場合では訓練回数に応じて平均ステップ数が増加し、最終的にはコースを周回し続けられるようになった。
- マイナスの報酬量を調整することで、学習モデルが大きく成長するようになったといえる。

参考文献

- 株式会社アフレル, 初版“ロボットで始める深層学習”，株式会社アフレル，2019年3月17日発行