平石研究室

深層強化学習によるライントレース制御に関する研究

S18129 武田翔太

1. はじめに

本研究では，深層強化学習を利用したライントレース制御について扱う．ライントレース制御ではセンサーによりラインの位置を認識し，どの程度，左右にずれている場合に，どの程度の速度で回転すれば，スムーズにラインをトレースできるかが問題となる．ライントレースに強化学習を適用する場合には，予めずれの大きさによって状態を分類しておく必要がある[1)]．それに対して，深層強化学習では，予め状態を分類しておく必要はなく，ニューラルネットワークの学習によって，自動的に状態が認識され，最適な行動が選択される．

本研究では，レゴマインドストームによる深層学習教材[2)]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．

2. シミュレーション環境

本研究ではPythonと株式会社アフレル製作の，「ロボットで始める深層学習」に記載されているプログラムを用いて環境の設定を行った．Python は 3.7.3 を用いた．コンピュータ上でライントレース制御の実験を行うため，教材記載のシミュレータ環境と実験用のプログラムを使用する．実験用のプログラムはロボットの動作とモデルの訓練を行い，JupyterLab で実行する．これらのプログラムを通信させることでシミュレータを動作させ，モデルの訓練及び訓練済みモデルのテストを行える．

3. モデルの訓練と訓練済みモデルのテスト

図1は，モデルの訓練のためのシミュレータの出力画面を示している．シミュレータ上にはロボットとそのロボットの視界，ロボットに追跡させる黒線が存在しており，黒線及びロボットの配置は設定ににより変更することができる．ロボットの視界はシミュレータ内の左上にも表示される．

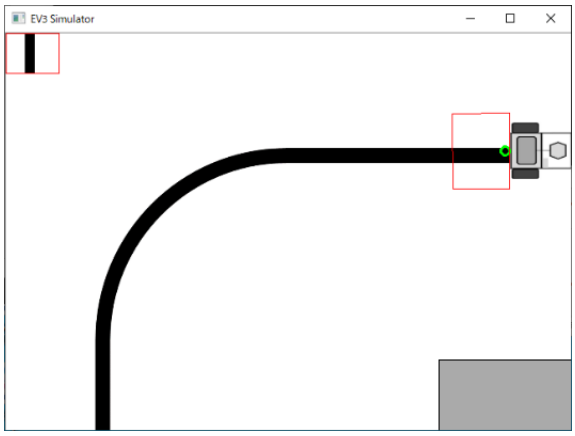

図1 モデル訓練用コース

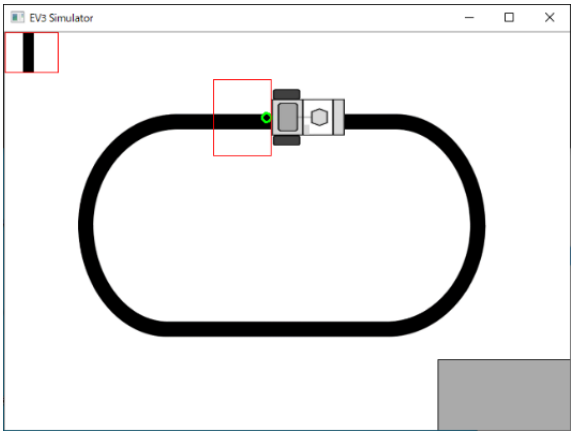

図2 訓練済みモデルテスト用コース

シミュレータ上ではロボットは前進し続け，前方の長方形状の範囲がロボットの視界であり，学習のための環境情報として利用する．ロボットの行動空間として左右にそれぞれ 2 段階の回転速度を用意しており，視界に捉えられた黒線の状態によって，どの回転速度が適切であるかをロボットは学習によって決定する．この時，黒線の重心座標が視界の中心からどれだけ離れているかを計算し，

近いほど高い報酬を与えるように設定している．そのため，より高い報酬が得られる回転速度に決定される．

また，視界内から完全に黒線が消えた場合には，ロボットは初期位置に戻され，モデルの訓練を継続する．ロボットの学習の効率のため図1のようにガーブ部分のみの 1/4 のコースでモデルの訓練を行う．モデルの訓練が終了したら，そのモデルをテストするために，図2に示したような全体のコースを利用して，訓練済みモデルのテストを実施する．

4．実験結果

表1　訓練ステップ数と実行ステップ数と報酬量

訓練	実行ステップ数		報酬量	
ステップ数	平均	標準偏差	平均	標準偏差
100	20.6	11.253	-1.14	0.234
200	13.8	1.166	0.81	0.137
300	10.4	0.490	0.23	0.198
400	17.8	1.939	1.02	0.153
500	37.4	14.541	0.49	0.364
600	27.8	2.482	1.03	0.119
700	237.8	180.671	0.93	0.306
800	22.4	2.059	1.17	0.317
900	32.2	13.991	1.89	0.158
1000	44	13.387	0.93	0.324
1100	33.4	21.453	5.36	8.320
1200	42.4	30.813	1.92	0.178
1300	45.8	29.721	1.69	0.192
1400	36	9.445	2.01	0.354
1500	25.6	6.184	1.51	0.242

本実験では，訓練することでモデルがどれだけ正確なライントレース制御を行えるようになるのかを調べるため，訓練するステップ数を 100 回毎増加させて 1500 回までデータを記録した．表1は実験結果を示した．実行ステップ数は視界内から黒線が消えるまでのステップ数であり，報酬量は1ステップあたりに得られた報酬の平均値である．それぞれ，5回実行した時の平均と標準偏差を示している．

予想では訓練ステップ数を多くすることで黒線を視界内にとらえ続ける実行ステップ数と報酬量は増大していくと考えていた．しかし，全てにおいて，周回し続けることができなかった．また，1 周には約 200 ステップが必要であり，一周することができたのは訓練ステップ数が 700 回の時のみであった．訓練ステップ数と実行ステップ数の相関係数は 0.075 であり，今回の実験において，訓練ステップ数と実行ステップ数の関連性は見られなかった．また，訓練ステップ数と報酬量の相関係数は 0.604 であり，僅かながら訓練ステップ数を増やすことでライントレースの正確性は上がっているものと考えられる．

5．おわりに

今回の利用した教材にしたがって実験を行なったが，モデルの訓練を行っても，ロボットを正確に周回させることができず，訓練ステップ数を増加させても，実行ステップ数に改善が見られなかった． 今回の実験において 700 回の訓練ステップ数において他に比べて，良い結果が得られているが，これは偶然優秀な記録を残したと考えられる．しかしながら，報酬量の改善はみられるため，訓練された環境おいて学習ができているものと考えられる．そのため，今後，訓練するコースを工夫し，訓練時にもテスト時と同じ環境を利用したり，失敗するカーブを追加で訓練するなどを行うことで，ロボットを正確に周回できるようにすることができるものと考えられる．

6．参考文献

1) 藤原滉司，平石広典，″二輪倒立ロボットのための強化学習による動作制御と行動選択″，情報処理学会第 75 回全国大会，Vol.2，pp.219-220，2013.3

2) 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