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．

これまでの研究において，深層学習教材 [2]にしたがってシミュレータ上での実験を行なったが，学習がうまくいかず，コースを周回し続けられる学習モデルの作成はできなかった．本研究ではコースを周回し続けることが可能な学習モデルの訓練方法について検証を行った．

2. モデルの訓練方法

訓練回数はこれまでの実験では 100 ステップから 1500 ステップまでを 100 ステップごとに行っていたが，強化学習にランダム性がある関係上，精度を高めるため，訓練回数を 1000 回から 5000 回までに増加させ，データは 1000 回ごとに取ることにした．また，学習モデルの訓練を効果的に行えるようにするため，訓練用コースを 3 種類用いて実験を行った．(図1)．

コース 1 はこれまでの実験で用いたものと同一であり，訓練回数を増加させるたことで生じる変化を調べるために利用した．コース 2 は作成した学習モデルのテストに使うコースを使用することでより効率の高い学習が行えるかを調べた．コース 3 は，これまでの実験でカーブでロボットが黒線から脱落してしまうことが特に多かったため,カーブを重点的に学習するために，コース 2 のロボットの初期位置をカーブの途中から開始したコースを使用した．

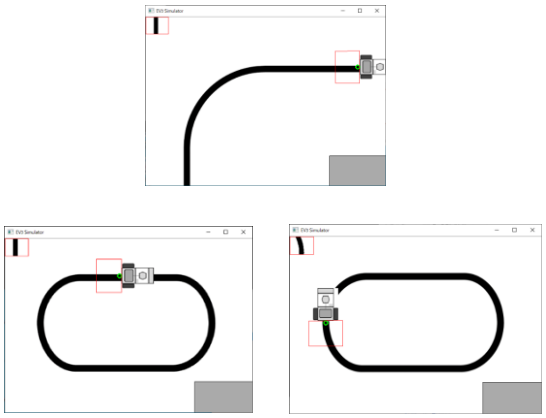

図1 コース 1(上) 教材での訓練用コース
コース2(左下) テスト用コースと同一
コース3(右下) カーブの途中から開始

3. 実験結果

表1 コース 1 から 3 の平均ステップ数

訓練回数	平均ステップ数		
	コース 1	コース 2	コース 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

表1はそれぞれのコースを利用して実験した結果をまとめたものであり，数値はロボットがコースを外れるまでのステップ数を示している．ただし，1000 ステップを実験の終了条件としている．表の数値は 20 回実行した際の平均のステップ数を示

している. コースを一周するのには約 200 ステップ必要である. そのため, ほとんどのケースにおいて一周させることができなかった. 平均ステップ数が最も高くなった訓練回数を見ると, コース 1 とコース 2 は 3000 回のときが最も高く, コース 3 は 1000 回のときが最も高くなった. コースを一周できたのはコース 1 の 1000 回と 3000 回の時のみで, どのコースと訓練回数も周回し続けられるものはなかった. また, 訓練回数を増加させることでステップ数が低くなるデータが見受けられた. この原因は訓練の途中でラインから少し離れた場合のマイナスの報酬を連続して獲得してしまい, ライントレースを続けることよりも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ためだと考えられる.

5. 報酬量を変化させての訓練

報酬量の問題を解決するため, 学習の際の報酬量を調整して実験を行った(表 2). 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に応じて報酬量を変化させている. この実験では距離が 15 以上ある際の報酬量を負の値から 0 に変更して実験を行った. ただし, ロボットが脱線する際に罰を与えるため, 距離が 40 以上の場合の報酬量は変更せず-6 とした. 報酬量を変えたことによる変化を調べるため, コース 2 を用いることとした.

表 2 ロボットの視界の中心と黒線の重心との距離 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
X≦40	-6	-6

表 3 報酬量を変化させた際の平均ステップ数

訓練回数	平均ステップ数		
	前半	後半	平均
1000	43.3	102.1	72.7
2000	56.5	162.7	109.6
3000	315.3	218.9	267.1
4000	636.0	227.3	431.7
5000	1000.0	628.0	814.0

表 3 は実験結果であり, 訓練を 5000 回まで 1000 回ずつ行った際の平均ステップ数をまとめたものである. これまで同様に全体で 20 回の実験をおこなったが, ここでは, 前半 10 回と後半 20 回に分けて結果をまとめた. この結果を見ると訓練回数が増加するにつれて, 平均ステップ数も増加していることが見て取れる. 前半では 5000 回まで訓練したところ, シミュレータの終了条件である 1000 ステップまで常に周回し続けられるようになった.

6. おわりに

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つづけられ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をするようになった.

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