深層強化学習によ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

研究の背景

- 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．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．

環境の構築

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う。

Pythonは3.7.3を用い、ライントレース制御をコンピュータ上で実行するためのシミュレータと動作・学習用のプログラムを動作させ、実験環境を構築する。

環境の構築

シミュレータはロボット・黒線・ロボットの視界から構成される。
ロボットの視界はロボットの前方の長方形の範囲で、画面の左上に表示されている。
黒線やロボットの配置を変えることや、学習する回数を変えることもできる。

ロボットは常に前進し、左右2段階ずつ回転することができる。

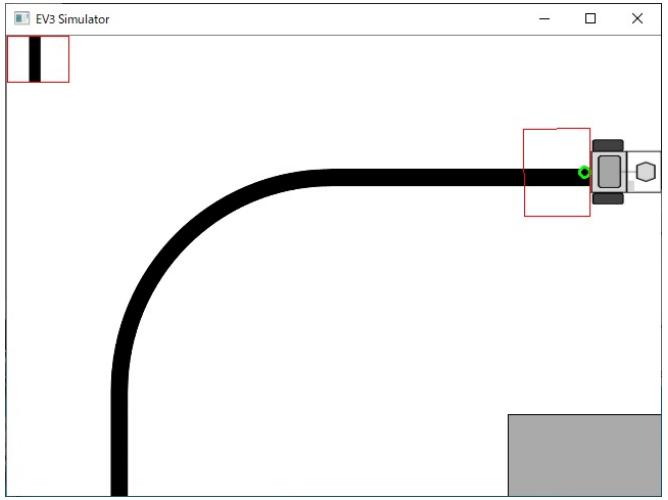


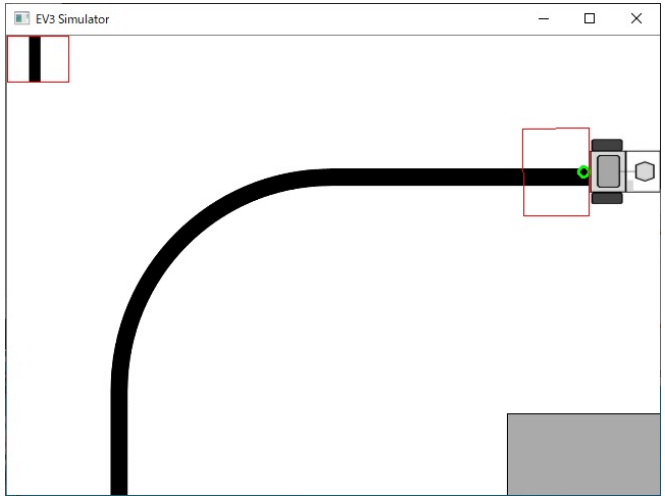
図　訓練中のシミュレータ

実験

学習をする際は図のような黒線の配置で規定のステップ数に達するまでロボットを動かし、モデルの学習を行う。

ロボットの視界の中心と黒線の重心が近いほど高い報酬が与えられるため、ロボットは黒線に沿って動くように学習する。

ロボットの視界内から黒線が
消えた場合は、ロボットを
初期位置に戻す。

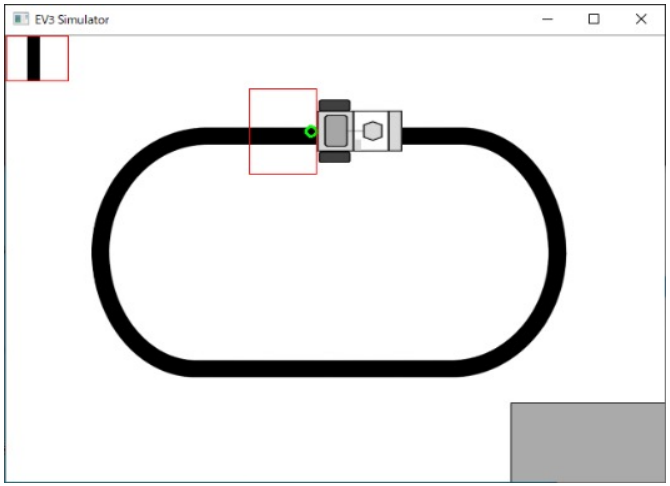


実験

学習モデルのテストは図のような楕円状の黒線の配置で行う。

学習の際と同様に視界の中心と黒線の重心が近いほど高い報酬が与えられる。この報酬の量を計測し、正確性の指標とする。

また、視界内から黒線が消えた際は
テストを終了し、それまでにかかった
ステップ数も記録する。



実験

学習するステップ数を１００回ずつ増やし、それぞれの学習モデルでテストを５回行った。

記録したステップ数は５回の平均と標準偏差を算出する。

報酬量はそのステップ数で割り、１ステップあたりの報酬量を算出してから５回の平均と標準偏差を算出し、グラフに纏める。

実験の結果

- 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、訓練量を増加させても実行ステップ数への影響がほぼ見られなかった。
- 訓練ステップ数と報酬量の相関係数は0.604であり、僅かながら訓練量を増加させることで獲得報酬量が増加した。

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

実験の結果

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．