仮想環境下における深層強化学習を用いた自動運転車でのラウンドアバウト走行検証

谷聖一 研究室　南 基大
Minami Motohiro

概要

近年，自動運転の実用化に向けた研究が盛んに行われている．その過程で種々の課題も見られる. その一つにラウンドアバウト（環状交差点）の通行がある. ラウンドアバウトは，人間による手動運転では多くの利点が指摘されているが, 現時点では自動運転には適しておらず, 研究課題となっている．報告者も欧米にて遭遇したが, 初遭遇でも問題なく使用でき,「手動運転で優れているのであれば, 人の神経回路網を模した深層学習でなら適応できるのでは」と考えた. ラウンドアバウトが本質的に自動運転に適しているかどうかを検証する端緒として, 本演習では統合開発環境の Unity にて仮想環境を構築し, 深層学習の一種である深層強化学習を用いた自動運転車にてラウンドアバウトの走行実験及び比較検証を行った.

1　はじめに

2020 年現在, 自動運転技術は waymo([1]) や tier4([2]) を代表とした各国企業にて研究開発が行われている. 自動運転には未だ課題は残っている. 複雑な環境では自立操作を諦め, 人に運転操作を渡してしまう. そのような環境の一つがラウンドアバウト ([3]) である．ラウンドアバウトは欧米では数多く導入されており, 手動運転では燃費や事故発生率などにおいて成果を上げている. しかし, 自動運転車によるラウンドアバウト通行は未だ未解決の問題の一つ ([4]) である．報告者も欧米にて遭遇したが, 初遭遇でも問題なく使用でき,「手動運転で優れているのであれば, 人の神経回路網を模した深層学習でなら適応できるのでは」と考えた．そこで, 深層強化学習を利用してラウンドアバウトを走行可能な自動運転モデルを試作した.

1.1　自動運転とは

自動運転とは, カメラやレーダ,GPS 等で周囲の環境を認識し, 人工知能によって運転手の補助もしくは自律運転を行うことである．米国自動車技術協会 (Society of Automotive Engineers) によって運転自動化レベルが全５段階で定義 ([5]) されている．そのうちレベル 3 からレベル 5 では自動車が運転主体であり人は非常時を除いて操縦する必要がないため, これらを自動運転と呼ぶ．現在, 世界各国の企業がレベル 4 自動運転の研究開発をおこなっている. 米アリゾナ州フェニックスの一部地域にて alphabet 傘下の waymo 社が無人運転によるタクシー業務の実用化 ([6]) を実現している．しかし, 未だ自動運転にも障害は多く存在している. 例として都市部などの密集地帯での運転, 複雑な交通システムでの運転が挙げられる．このような環境下では自動運転車は自立操作を諦め, 人に運転操作を渡してしまう場合が多々発生する.

1.2　ラウンドアバウトとは

ラウンドアバウト (roundabouts) とは, 交通工学研究会のラウンドアバウトの計画・設計ガイドによると,「環道交通流に優先権があり，かつ環道交通流は信号機や一時停止などにより中断されない，円形の平面交差部の一方通行制御方式」([7]) である．アメリカでは 2010 年時点で設置数が 2000 を超えており ([8]), 日本でも 2015 年時点で 32 都道府県に 140 箇所程度存在 ([9]) している. 手動運転において燃費, 事故発生率 ([10, 11]) の統計的優位性が報告されている.

1.3　強化学習とは

強化学習 (reinforcement learning) とは，環境との試行錯誤による相互作用を通して適切な行動戦略を獲得するタイプの機械学習である ([12]). 自律ロボットに関する研究が活発に行なわれている. 環境との相互作用を通して自律的にデータを集めて学習を行なう強化学習は大量のデータを用意する必要が無い学習制御アルゴリズムとして注目されている.

強化学習では, 学習の主体者であるエージェント (agent) は環境 (environment) の状態 (state) を観測し, それに応じ行動する．行動によって環境が変化し, エージェントは差から報酬を受け取る. この試行錯誤を繰り返すことで最終的にエージェントは最も大きい報酬を得られるように方策 (policy) を変化させる．問題の規模や特性によっては現実世界では試行錯誤が困難であることが多い. その為, 現実世界を模した仮想環境を構築し, その上でエージェントに学習させる手法が取られる．方策

の関数にニューラルネットワークを利用したものを深層強化学習といい, 研究が盛んに行われている.

1.4 背景

自動運転の学習用データセットは複数の研究施設や企業により公開されている. しかし, ラウンドアバウトは交差点の一種でしかないため, 走行データの数が少なく, 学習に耐え得るものではない. そのため, 大量のデータがなくとも自律的に学習する強化学習を用いることでラウンドアバウトを走行できる自動運転エージェントを作成することにした.

1.5 演習目的・内容

仮想環境上に「学習用環境」「ラウンドアバウト環境」「十字交差点環境」計 3 種の学習・評価用環境と自動運転車を構築する. 構築した走行環境にて自動運転車が深層強化学習を用いて学習を行い, ラウンドアバウトを走行する運転モデルを作成する.

2 演習方法

2.1 実行・開発環境

開発環境は以下の通りである.

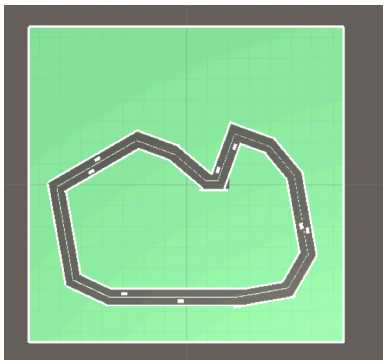
- 使用言語
  - Python3.5
  - ソフトウェア
    - Unity[13]
  - ライブラリ
    - ML-Agents
    - TensorFlow
    - TensorFlow-GPU

2.2 学習・評価環境の構築

学習に用いる仮想環境を構築する. 本演習ではゲーム・シミュレータ開発によく使われ, モデリングや物理演算が容易な統合開発環境の「Unity」を用いた. 学習・評価を行うために走行環境を用意した後に自動運転車を構築した. 走行環境, 自動運転車は実際の日本交通システムと同スケールにて構築した.

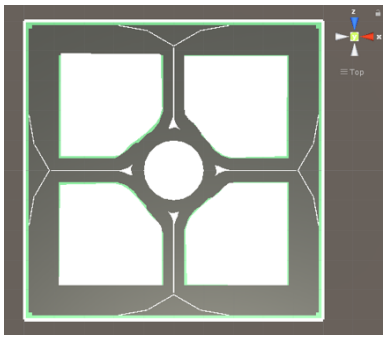
2.3 走行環境の構築

走行環境には「学習環境 S:サーキット」(以下, 学習環境 S と呼ぶ)「評価環境 R:ラウンドアバウト」(以下, 学習環境 R と呼ぶ),「評価環境 X:十字交差点」(以下, 学習環境 X と呼ぶ) の 3 種類を用意した.



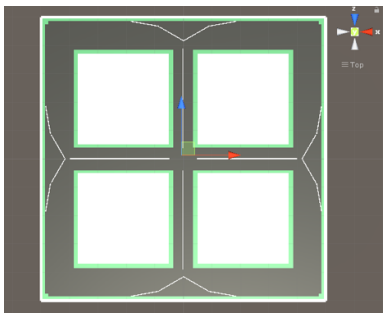
大きさ	環境全体	150m × 150m
道幅	道路	9m

図 1: 学習環境 S:サーキット



大きさ	環境全体	100m × 100m
道幅	環状道路	7-9m
	外周道路	9m

図 2: 評価環境 R:ラウンドアバウト



大きさ	環境全体	100m × 100m
道幅	中央十字路	7m
	外周道路	9m

図 3: 評価環境 X:十字交差点

2.4 自動運転車の構築

車体性能

自動運転車は一般的な普通自動車を参考に構築した. 実装の都合上, 加速・減速時の速度変化は一定量であり, 旋回時は瞬間的に 3 車体が回転する . 一秒間に 30 回の周囲認識・動作決定を行い, 運転は「加速・等速・原則」「左折・直進・右折」の 2 種類のグループからそれぞれ 1 つづつ選択する.

最高速度	60km/hour
旋回速度	50 ° /s
加速度	20m/s

図 4: 車体性能

センサー

人は運転する際に周りの環境を目視にて周囲を認識し運転するが. 自動運転車の場合はセンサーやカメラを用いて周囲を認識し運転する. 特にセンサーに LIDAR(Light Detection and Ranging) を用いた周囲認識が主流な為, 本演習では LiDAR を簡略化して実装した. 22.5 間隔に合計　16 本のセンサーが車体中央を基点に円形に配置され, 各センサーは壁, 白線, 他車までの距離・衝突接点の角度を検知する.

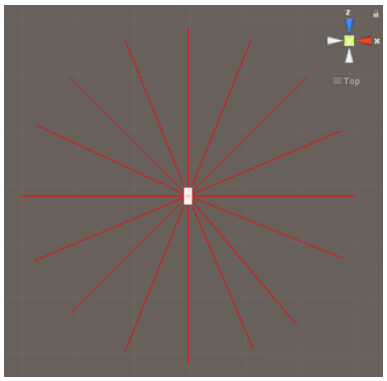


図 5: 自動運転車及びセンサー

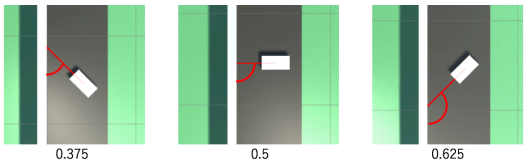
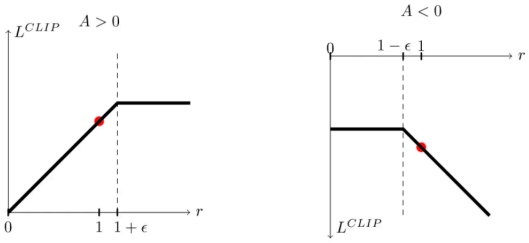


図 6: センサー角度は壁との接点から算出した

2.5 自動運転モデルの学習

構築した仮想環境上にて深層強化学習を用いた学習を行った. 予備実験にて学習に評価環境 R, 評価環境 X を用いたところ, 学習が安定しなかった. 学習過程を目視で確認したところ, 交差点にて衝突が多く発生して学習が停滞していると判断し, 学習環境 S を構築した. 学習は 800,000step までは学習環境 X を用いた基本運転の学習を行い, その後に 1,000,000step まで評価環境 X・R を用いた発展運転の学習の 2 つのステップに分かれている. そのうちの「評価環境内運転の学習」にて「十字交差点とラウンドアバウトを同時に学習」「十字交差点とラウンドアバウトを同時に学習した後にラウンドアバウトのみで学習」「十字交差点とラウンドアバウトを同時に学習した後に十字交差点のみで学習」「十字交差点の後にラウンドアバウトで学習」「ラウンドアバウトの後に十字交差点で学習」の合計 5 種類のモデルを作成した. また, 深層強化学習の勾配関数には Proximal Policy Optimization Algorithms (PPO, 2017)[14] を利用した. PPO は step 関数を用いることでパラメーターの急激な変化を防いでおり, 学習が安定しやすい特徴を持つ.



$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\right]$$

図 7: PPO

本演習では，375 次元の中間層を 3 層持つ PPO モデルを用いる.

2.6 学習報酬設定

強化学習において, 報酬は開発者が決める必要があるパラメーターである. 基本的にエージェントに対しどのような行動をさせたいかを考え, 設定する必要がある. 例として, 事前実験にて図 8 のような報酬にて学習を行った. 本来想定した動きは「できるだけ高速にて移動し, 左車線にとどまるようにする」というものだった.

条件	値
速度 (0-最高速度で正規化)	+0~1
左車線に居る (右 67.5-112.5 度, 距離 2m 以内に白線が存在)	+1
右車線にいる (左 247.5-297.5 度, 距離 2m 以内に白線が存在)	無し
建物・車に衝突	-500

図 8: 誤った学習をした報酬設定

しかし, 実際の動きとしては「スタート地点から一切動かず, 左車線にいる報酬のみを獲得し, 衝突が発生しないようにする」という開発者の意図とは違う学習になった. このように, 報酬は様々な値を実際に学習を行いながら試行錯誤で設定する必要がある.

本演習では事前実験にて最も成績の良かった図 9 の報酬設定にて学習を行った.

条件	値
速度 (0-最高速度で正規化)	+0~1
左車線に居る (右 67.5-112.5 度, 距離 2m 以内に白線が存在)	+0.5
右車線にいる (左 247.5-297.5 度, 距離 2m 以内に白線が存在)	-0.5
建物・車に衝突	-500

図 9: 本演習にて使用した報酬設定

同じハイパーパラメータでも局所解に陥ってしまう事が有るため. 累計報酬の推移を確認しながら学習を行う. 本演習では事前実験を参考に 1,000,000step まで学習を行った. 図 10 が累計報酬の推移である.

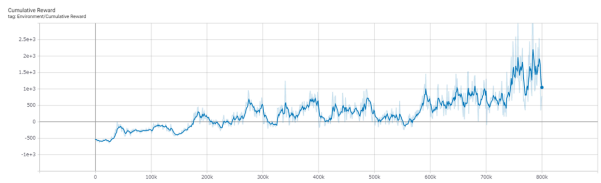


図 10: 累計報酬の推移

3 結果考察

構築した自動運転モデルを用いて, 評価環境 R, 評価環境 X にそれぞれ 1 時間走行し, 1 分毎の平均を評価値とする. 評価環境に設置する車両数は 8 両と 16 両の 2 種類を用意し, 混雑時と比較する. また, 評価基準は以下の 2 つを用意した.

Clash:衝突発生数

交差点内にて建物, 車に衝突した車数. 各モデルがどれだけ安全に交差点を通行できるかの評価指標を 設定した.

Flow:車両通過数

交差点を通過した車数. 交差点内で衝突した車両は含まれない. 各モデルがどれだけ潤滑に交差点を通行できるかの評価指標を意図して設定した.

3.1 評価結果

評価結果を以下に示す.

車両数	MIX		MIX→C		MIX→R		C→R		R→C	
	R	X	R	X	R	X	R	X	R	X
8	8.10	13.94	6.89	15.45	7.00	13.88	9.45	16.69	7.49	16.72
16	18.38	46.32	17.44	43.80	14.86	51.32	19.26	48.87	17.37	41.52

図 11: 評価結果:Clash

車両数	MIX		MIX→C		MIX→R		C→R		R→C	
	R	X	R	X	R	X	R	X	R	X
8	49.10	39.70	43.11	41.65	48.73	42.11	41.30	44.30	46.96	40.72
16	57.20	37.80	53.84	36.57	62.23	34.97	43.17	44.37	51.70	42.45

図 12: 評価結果:Flow

3.2 考察

十字交差点ではラウンドアバウトに比べて Clash が平均 2.32 倍発生した. ラウンドアバウトは深層強化学習による自動運転にとって比較的容易に走行可能と推察される. また, 車両数を 8 から 16 にした際, ラウンドアバウトでは平均 2.25 倍, 十字交差点では平均 3.02 倍 Clash が発生した. ラウンドアバウトは十字交差点に比べ車両密集時に走行しやすいと推察される. しかし,Flow に関しては有意差を確認できなかった. 通過する速度を見るには Clash による影響が大きすぎたと推察される.

3.3 今後の課題

交差点内の移動速度を確認するために Flow を用意したが, 評価に使える指標たり得なかった. 交差点内に侵入してから出るまでの平均時間など, 違う評価指標が必要と考えられる.

本演習では提案者が慣れている Unity を用いた仮想環境で構築を行ったが, あまり現実に即しているとは言いずらい状況である. 今後, 自動運転システム OSS の「Autoware」[15] を用いて, より現実に近い環境で実験を行う予定である.

参考文献

[1] Waymo が自動運転車の次世代技術を Jaguar I-Pace で試験中 https://jp.techcrunch.com/2020/03/09/2020-03-06-inside-the-next-gen-tech-on-waymos-self-driving-jaguar-i-pace/

[2] https://tier4.jp/

[3] 国土交通省, ラウンドアバウトの現状報告書, https://www.mlit.go.jp/road/ir/ir-council/roundabout/pdf01/4.pdf

[4] Machine Learning Techniques for Undertaking Roundabouts in Autonomous Driving https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6566321/

[5] JASO,"自動車運転自動化システムのレベル分類及び定義",2018, https://www.jsae.or.jp/08std/data/DrivingAutomation/jaso_

[6] TABILAB, 運転手のいないタクシー「Waymo」がアリゾナ州フェニックスで本格始動,

https://tabi-labo.com/293151/wt-waymo

[7] ラウンドアバウトの計画・設計ガイド, 交通工学研究会,2009

[8] Current Roundabout Practice in the United State Kittelson Associates.Inc,International Roundabout Design and Capacity Seminar 6th International Symposium on Highway Capacity Stockholm, Sweden, 2010

[9] ラウンドアバウトの現状報告書, 国土交通省,2015

[10] Hang Cao.Máté Zöldy,An Investigation of Autonomous Vehicle Roundabout Situation,2019

[11] Roundabouts: An Informational Guide Second Edition”, NCHRP Report 672 p136,FHWA,2010. https://nacto.org/docs/usdg/nchrprpt672.pdf

[12] Sutton, R.S., Barto, A.G.:“Reinforcement learning: An introduction, a bradford book”, MIT Press, 1998. (邦訳 三上, 皆川: “強化学習”, 森北出版, 2000.

[13] https://unity.com/ja

[14] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov,”Proximal Policy Optimization Algorithms”,2017.

[15] https://github.com/autowarefoundation/autoware