

卒業論文 2019 年度 (令和 1 年)

目的満足度に応じたモビリティ全体制御

慶應義塾大学環境情報学部環境情報学科
島津翔太

目的満足度に応じたモビリティ全体制御

自動運転技術の発展により、我々人間は車を常時コントロールする必要がなくなってきた。例えば、車線の自動変更や物体検知による自動ブレーキなどは、今まで人間が行っていた操作をコンピューターによって行っている。今後、自動運転技術はより高度になり、近い将来人間がハンドルを握らなくともコンピューターによる制御のみで目的地まで到達できるレベルまで発展する可能性がある。そのような自動運転車が普及した社会において、自動車各個が個別に行動すると様々な問題が発生する。

第一に、特定の経路の混雑があげられる。多数の自動運転車が個別に経路を選択した場合、特定の道が混雑する問題が発生する。第二に、自動運転による使用用途の変化への順応である。人間による操作が一切行われない完全自動運転が実現する事はすなわち無人運転が可能であることを意味する。無人運転が可能となった場合、自動車の所有者が目的地についた後に駐車する必要はなく、駐車時間中に自動車を使いたい人の元へ迎えに行く無人のヒッチハイクのような行為が可能となる。このような使用用途の変化に対応するには、単に目的地までのルートを選択するだけではなく目的地に到達した後の行動も自動運転車が決定する必要がある。

そこで、本研究では強化学習を用いて自動車の利用者の目的に合わせた経路選択の実験を行った。実験では、緊急のミーティングに向かっている、空腹を感じているなどの利用者の目的や状態などのパターンを予め想定した。そのパターンに対して利用者の目的を達成するルートを選択した場合に強化学習に報酬値を与え、目的を満たせないルートを選択した場合は罰則値を与える。これを繰り返すことにより、強化学習アルゴリズムがより目的を達成するような経路選択を行えるようになると考えられる。

これらの実験を通して、本研究では正解データなし自律して意思決定を行うアルゴリズムの一種である強化学習を用いたアプローチが人間の目的を満たす意思決定に有効であること、モビリティの経路選択において実用的であることを示す事を示す。

キーワード:

1. 機械学習, 2. 深層強化学習, 3. モビリティ, 4. シェアリングエコノミー

慶應義塾大学環境情報学部環境情報学科
島津翔太

Mobility for Human Satisfaction

With the development of autonomous driving technology, humans no longer need to constantly control cars. For example, automatic lane changes and automatic braking based on object detection are done by computers in the same way that humans have done so far. In the future, automatic driving technology will become more advanced, and there is a possibility that in the near future, it will be developed to a level where a person can reach a destination only by computer control without grasping the steering wheel. In a society where such self-driving vehicles are widespread, various problems arise when individual vehicles act individually.

First, there is congestion on specific routes. If a large number of autonomous vehicles select their own routes individually, there will be a problem of congestion on specific roads. The second is to adapt to changes in usage by automatic operation. The realization of fully automatic operation without any human intervention means that unmanned operation is possible. When unmanned driving becomes possible, it is not necessary to park a car after the car owner arrives at the destination, and an act such as an unmanned hitchhiking to pick up a person who wants to use the car during parking time becomes possible. In order to respond to such changes in usage, it is necessary for the self-driving vehicle not only to select the route to the destination but also to determine the actions to take after reaching the destination.

In this study, therefore, we use reinforcement learning to experiment with route selection according to the purpose of vehicle users. In the experiment, we assumed the user's purpose and condition such as going to an emergency meeting or feeling hungry. If the user selects a route that achieves the user's purpose for that pattern, we give a reward value to reinforcement learning. If the user selects a route that does not meet the user's purpose, we give a penalty value. By repeating this process, the reinforcement learning algorithm may be able to select a route that achieves a better goal.

Through these experiments, we hope to show that an approach using reinforcement learning, which is a kind of algorithm that makes independent decisions without correct data, is effective for making decisions that satisfy human objectives and is practical for path selection of mobility.

Keywords :

1. Machine Learning, 2. Reinforcement learning, 3. Mobility, 4. Sharing Economy

Keio University Faculty of Environment and Information Studies
Shota Shimazu

目次

| | | |
|-------|-----------------------------|----|
| 第1章 | 序論 | 1 |
| 1.1 | 本研究の目的 | 1 |
| 1.2 | 本論文の構成 | 1 |
| 第2章 | 背景 | 3 |
| 2.0.1 | 自動運転 | 3 |
| 2.0.2 | Mobility as a Service | 3 |
| 2.0.3 | シェアリングエコノミー | 4 |
| 2.0.4 | 機械学習 | 4 |
| 第3章 | 本研究における問題定義と仮説 | 5 |
| 3.1 | 最短経路問題 | 5 |
| 3.2 | 満足度最適化 | 6 |
| 3.3 | 本研究における問題定義 | 6 |
| 3.3.1 | 強化学習によるルート選択 | 6 |
| 3.4 | 仮説 | 7 |
| 第4章 | 提案手法 | 8 |
| 4.1 | 概要 | 8 |
| 第5章 | 要素技術 | 9 |
| 5.1 | 機械学習 | 9 |
| 5.1.1 | 深層学習 | 9 |
| 5.1.2 | 強化学習 | 10 |
| 5.1.3 | 深層強化学習 | 10 |
| 5.2 | Q学習 | 11 |
| 5.3 | Deep Q Neural Network | 11 |
| 第6章 | 実験 | 12 |
| 6.1 | 概要 | 12 |
| 6.2 | 構成 | 12 |
| 6.3 | Deep Q Neural Network Model | 13 |
| 6.4 | Data Serializer | 13 |
| 6.5 | 可視化地図アプリケーション | 13 |

| | | |
|--------------|-----------------------------|-----------|
| 6.6 | Docker | 13 |
| 6.7 | 本研究で用いる環境定義アルゴリズム | 14 |
| 6.8 | 環境定義クラス | 14 |
| 6.9 | 本実験における前提 | 15 |
| 第 7 章 | 評価 | 17 |
| 7.1 | 評価内容 | 17 |
| 第 8 章 | 結論 | 18 |
| 8.1 | 本研究のまとめ | 18 |
| 8.2 | 本研究の課題 | 18 |
| 8.2.1 | 想定環境の貧弱さ | 18 |
| 8.2.2 | 学習器の連携 | 18 |
| 付 録 A | APPENDIX | 20 |
| A.1 | 本論文で用いる機械学習関連の略語 | 20 |
| | 謝辞 | 21 |

図 目 次

| | | |
|-----|-------------------------------|----|
| 3.1 | 最短経路問題 | 5 |
| 3.2 | 強化学習による最短経路検索の結果 | 6 |
| 5.1 | パーセプトロン | 10 |
| 5.2 | 強化学習の概念図 | 11 |
| 6.1 | システムの関係図 | 12 |
| 8.1 | 複数エージェントで環境を共有する概念図 | 19 |

表 目 次

| | | |
|-----|-----------------------------------|----|
| 6.1 | アプリケーション作成に用いたソフトウェアと詳細 | 13 |
| 6.2 | 本実験で適用する条件 | 16 |

第1章 序論

本章では本研究の動機，本研究の概要を示す．

Abstract からポート 自動運転技術の発展により，我々人間は車を常時コントロールする必要がなくなってきた．例えば，車線の自動変更や物体検知による自動ブレーキなどは，今まで人間が行っていた操作をコンピューターによって行っている．今後，自動運転技術はより高度になり，近い将来人間がハンドルを握らなくともコンピューターによる制御のみで目的地まで到達できるレベルまで発展する可能性がある．そのような自動運転車が普及した社会において，自動車各個が個別に行動すると様々な問題が発生する．

第一に，特定の経路の混雑があげられる．多数の自動運転車が個別に経路を選択した場合，特定の道が混雑する問題が発生する．第二に，自動運転による使用用途の変化への順応である．人間による操作が一切行われない完全自動運転が実現する事はすなわち無人運転が可能であることを意味する．無人運転が可能となった場合，自動車の保有者が目的地についた後に駐車する必要はなく，駐車時間中に自動車を使いたい人の元へ迎えに行く無人のヒッチハイクのような行為が可能となる．このような使用用途の変化に対応するには，単に目的地までのルートを選択するだけではなく目的地に到達した後の行動も自動運転車が決定する必要がある．

1.1 本研究の目的

将来の完全自動運転車が実現するにあたり，**加筆・修正** 本研究では，強化学習を用いた経路選択を試行し自動車の利用者の移動目的をより多く達成できる強化学習モデルの構築を目指す．現状，自動運転車に代表されるモビリティは個別に行動しており，ルートの選択に置いて人間が与えた目的地への最短経路を選択している．この場合，一定時間内に同一方面の目的地を多数設定された場合に特定のルートの混雑が予測される．また，レベル5の完全自動運転が実現された場合，自動車を使用していない期間は自動で他者に貸し出すなどの利用方法の変化も見込まれる．

本研究では，機械学習モデルの一つである深層強化学習を用いて，経路選択の実験を行い，自律意思決定型の機械学習アルゴリズムが自動運転車において活用可能であることを示したい．

1.2 本論文の構成

本論文における以降の構成は次の通りである．

2 章では，背景を述べる． 3 章では，本研究における問題の定義と，解決するための要件の整理を行う． 4 章では，本研究の提案手法を述べる． 6 章では，4 章で述べたシステムの実装について述べる． 7 章では，3 章で求められた課題に対しての評価を行い，考察する． 8 章では，本研究のまとめと今後の課題についてまとめる．

第2章 背景

本章では、本研究の背景を示す。

2.0.1 自動運転

車における自動運転は1980年代から研究されてきた。例えば、欧州で1987年から1995年に行われたEUREKA プロメテウス計画¹では高速道路における車線の追従や車線の変更などの自動運転の基礎技術が研究された。現在では、これらの機能は市販の自家用車にも運転を支援する機能として搭載されている。また、高速道路など限定した場所であれば人間による介入が不要な一部自動運転が可能となっているものもある。今後、将来自動運転技術はより人間による操作を少なくし、首相官邸ホームページ「官民 ITS 構想・ロードマップ 2019」²に定義されたレベル5の完全な自動運転技術が完成する可能性がある。

2.0.2 Mobility as a Service

日本に置いて、車や鉄道などの交通は高度経済成長期以降、道路網や路線網の拡大も合わせて急速に普及が進み、旅客・貨物共に主たる移動手段となった。

しかし、近年、交通は単なる移動手段としてだけではなく、移動や移動に付随する付加価値や自己所有の車を自ら運転すると行った従来の使い方からの変化が求められるようになってきた。

これに対して移動をサービスとして提供するという考え方があり、Mobility as a Serviceの略称MaaSと呼ばれている。現在MaaSサービスとしては米Uberなどに代表される個人所有の車を配車するサービスや、自動車を不特定多数の利用者で共有するカーシェアリングサービスなどがある。

人間による操作を必要としないレベル5の自動運転が実現すると、ハンドルを握る必要がないため従来のように自動車において移動時間中に運転に拘束されることがない。また、移動経路も人間が考えることなく目的地まで到達する。このようになると、単に自動車そのものを共有するサービスだけではなく移動経路の選択や移動時間の活用などのサービスとして提供する事が必要になると予想される。

¹<https://www.eurekanetwork.org/project/id/45>

²<https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20190607/siryou9.pdf>

2.0.3 シェアリングエコノミー

シェアリングエコノミー [1] ないしは共有経済とはモノやサービスを特定の個人で所有するのではなく複数人で共有する社会関係もしくは社会システムである。古くは GNU プロジェクトなどのオープンソースソフトウェアがそれに当たると考えられている。スマートフォンの普及により人々は常時インターネットに接続するようになった, これにより Uber¹ などの個人所有の車の配車サービスや AirBnb² などの所有する不動産を一定期間旅館のような形で貸し出すと行ったサービスが登場し, 社会的に大きな影響をもたらしている。将来, 人間が介在することのない自動運転が可能になると自動車そのもののハードウェアをシェアするだけではなく自動車を使ったサービスのシェアが加速すると思われる。

2.0.4 機械学習

機械学習とはコンピューターが自動的にパターンを学習し人間による指示がなくとも課題を実行する事である。モバイルコンピューターの性能が飛躍的に向上し画像の認識や機械翻訳などの多くの分野で応用されるようになってきた。これらの技術は日々進化を重ねており画像分類の分野においては 1 年で認識率を 20% 向上したり, 学習にかかる時間を 50% ほど短縮するアルゴリズムが考案されたりと飛躍的に向上してる。また, 機械学習アルゴリズムは意思決定を伴う処理にも活用され, AlphaGo [2] に代表されるように囲碁などのボードゲーム分野において人間をも上回る成果を出している。加えて, 意思決定型の機械学習はボードゲーム意外にも応用されつつあり, リスティング広告の最適化など利用者である人間にとっての利益が最大化するような制御に応用されつつある。

¹<https://www.uber.com/jp/ja/>

²<https://www.airbnb.jp>

第3章 本研究における問題定義と仮説

本章では、第2章で述べた背景から、現状の自動運転システムが目的としている制御方法の問題点を述べる。

3.1 最短経路問題

広く一般的に自動運転車が普及すると、人間による非合理的なブレーキなどの操作がなくなり渋滞が解消すると言われている。しかし、渋滞は車線あたりの交通量に比例し、一般的に自動車の走行ルートを決定的する場合、現在地から目的地までの最短ルートを選択するため交通需要の高い経路の交通量は変わらず、自動運転による渋滞解消の因果関係には疑問が残る。少なくとも同時時間帯に同一地点付近の目的地を設定した車が大量にいた場合に特定のルートが混雑が発生することは避けられないと考えられる。現状のカーナビゲーションシステムなどに搭載されているような渋滞回避機能であっても、最短経路の周辺の道路に回避され最短経路を選択するには早いもの順になる状態になる。



図 3.1: 最短経路問題

3.2 満足度最適化

将来のモビリティのサービス化を進めるにあたり、目的地までの最短経路の走行では利用者の目的を満たせず満足度を高められない可能性がある。例えば、

3.3 本研究における問題定義

現状の強化学習の問題点を列挙し、整理する。

3.3.1 強化学習によるルート選択

経路選択に強化学習を応用した事例 [3] はすでに存在する。例えば、下図のような単純な正方形の升目上の任意の座標にスタート地点とゴール地点を設定し、最短経路を学習するという物である。しかし、実際の道路網は迷路と比較しても非常に複雑であり沿線との接続や経由地などの要素を考慮す

[ここは修正・加筆必要] 強化学習を応用した場合の精度は未知数である。

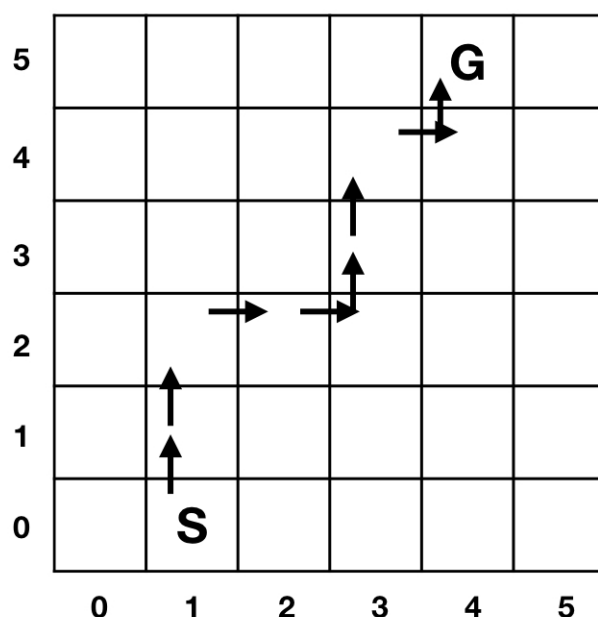


図 3.2: 強化学習による最短経路検索の結果

3.4 仮説

Deep Q Nueral Network により学習を繰り返すことで, シミュレーターないでのではないでの利用者に最も最適化された経路選択を行うようになると思った.

第4章 提案手法

本章では提案手法について述べる.

4.1 概要

本研究では, 人間の状態とモビリティの経路を再現するシミュレータープログラムを開発し実験を行う. なお, シミュレータープログラムの作成にあたっては以下について仮定を行った.

- シミュレーター上での自動車は全てレベル5の完全自動運転ができる物とする
- 歩行者や, 自動運転非対応者などは考慮していない
- シミュレーターでは予め決められた幹線道路及, 環状線などのバイパス道路のみを考慮し, それ以外の道に関しては考慮しない
- 人間の目的を正確に推定するコンピューターシステムもしくはセンサーなどは実在しているものとする.

第5章 要素技術

本章では本研究に用いる手法に関連する要素技術について述べる

5.1 機械学習

機械学習とは広義には、コンピューターが自動的にパターンを学習し人間による明示的な命令がなくとも特定の課題を自動で実行する技術又はアルゴリズムのことである。主に、正解データを与えることによってパターンを学習する教師あり学習、データのまとまりや相関を求める教師なし学習と繰り返し反復することで価値が最大化するように学習を行う強化学習に分類される。画像処理や自然言語処理など様々な分野に応用が進んでおり、画像の検知や自動翻訳など機械学習によって自動化されているものも多い。

5.1.1 深層学習

深層学習とは脳が持つ脳神経系のニューロンをソフトウェアで再現した人工ニューラルネット (ANN) を持つ機械学習アルゴリズムの一つである。人間の脳を模したパーセプトロン [4] による深層学習自体は 1957 年から提唱されていた。しかし、4 層以上のパーセプトロンでは過学習や勾配消失問題が発生しやすく計算コストも大きいためあまり普及しなかった。しかし、安価なコンピューターでも計算速度が飛躍的に向上したことや勾配消失問題を防止する手法が考案されたことなどから再び注目を集めるようになり、様々な課題に特化した DNN が考案されている。

例えば、Convolutional Neural Network (CNN) は画像の特徴量抽出に長けており画像認識分野で大きな成果をあげている。Long short-term memory (LSTM) は自動翻訳やテキスト解析など自然言語処理分野で活用されており、現在では言語だけではなく音楽などにも活用範囲が広がっている。また、既存の機械学習の手法に深層学習を組み込んだ事例も散見される。

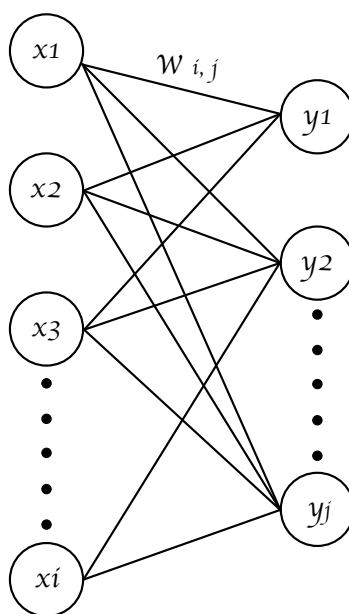


図 5.1: パーセプトロン

5.1.2 強化学習

強化学習 (Reinforcement learning) はエージェントと呼ばれる行動主体が現在の状態を観測し価値が最大化する行動を繰り返し選択することにより利益が最大になる行動を学習する機械学習手法の一種である。強化学習では行動主体であるエージェントと環境を定義する状態と行動した結果による変化, 報酬が定義される。初期的な強化学習にはマルコフ決定過程や Q 学習というものがある。

5.1.3 深層強化学習

深層強化学習とは, 強化学習に深層学習を組み合わせた機械学習アルゴリズムである。強化学習はエージェントとエージェントが動作する環境を定義し, 定義された環境下でエージェントへの報酬が最大化するように学習は行う。

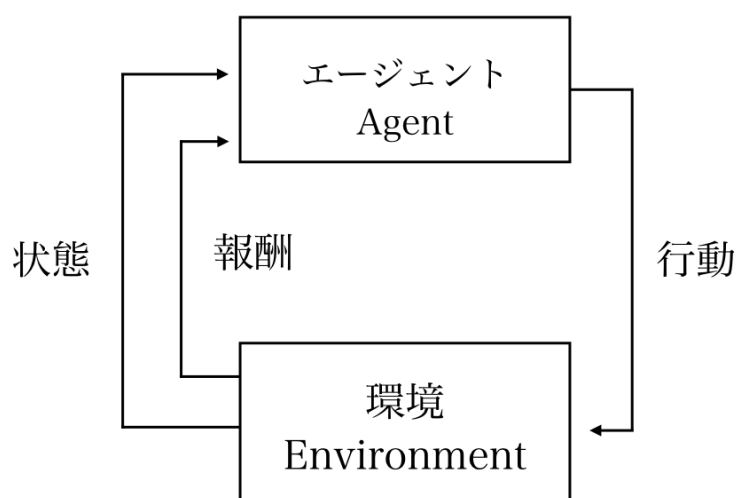


図 5.2: 強化学習の概念図

5.2 Q 学習

Q 学習とは初期に考案された Reinforcement Learning アルゴリズムの一種であり, 強化学習と呼ばれす. Q 学習による試行は以下のように定義する.

$$Q(s, a) \approx R(s, a) + \gamma \max_{a'} E[Q(s', a')] \quad (5.1)$$

5.3 Deep Q Neural Network

Deep Q Neural Network(DQN [5]) は Q 学習という強化学習における古典的なアルゴリズムを深層強化学習に応用したものである. Q 学習とは強化学習の一種である. Q 学習では実行するルールに対して Q 値という値を持たせる.

DQN は深層強化学習とも呼ばれる.

第6章 実験

本章では提案手法の実験について述べる.

6.1 概要

本研究では, 以下に述べるコンピュータプログラムを実装し, モビリティや人の動きをコンピュータ上で再現する. 作成するプログラムは 第一は, 強化学習モデルを生成する機械学習プログラムと

6.2 構成

この実験での各プログラムやミドルウェアの関係性をいかに示す.

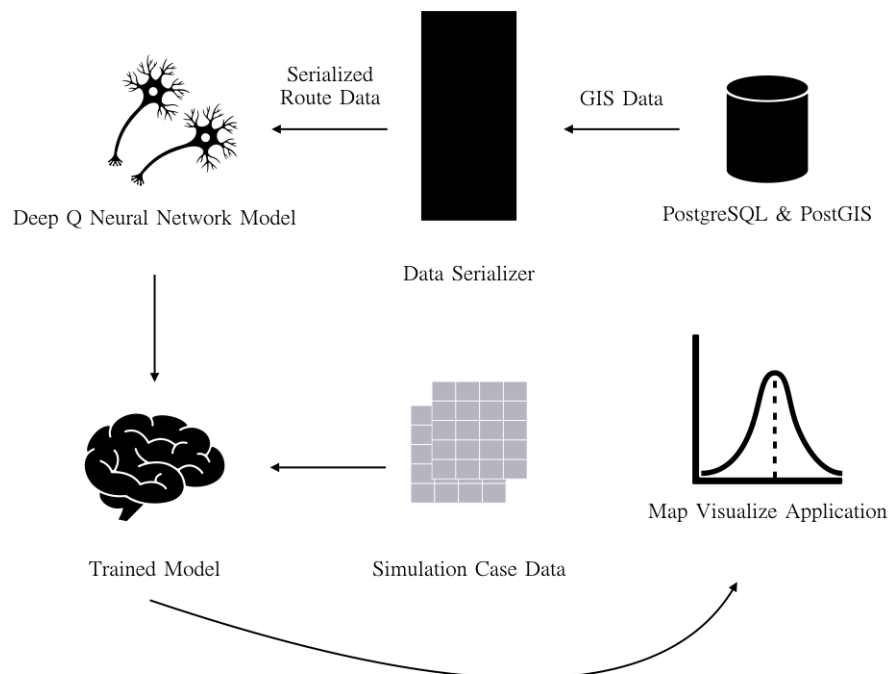


図 6.1: システムの関係図

6.3 Deep Q Neural Network Model

このプログラムは Deep Q Neural Network による強化学習を行うプログラムである。実装言語は Python である。

6.4 Data Serializer

Data Serializer は PostgreSQL サーバー上に記録された経路などの地理データを上述した Deep Q Neural Network に学習させるためのデータ変換を行うプログラムである。開発言語は Python であり、一部データベースとの接続部分に C++ を用いた。

6.5 可視化地図アプリケーション

モビリティの動きを地図上に可視化するアプリケーションの作成を行った。このアプリケーションは Deep Q Neural Network の学習済モデルが下した意思決定つまり経路を地図上にアイコンとして再現し、乗車している人間に関しての情報を統合して閲覧できるようにしたものである。

表 6.1: アプリケーション作成に用いたソフトウェアと詳細

| 使用対象 | 使用言語 | フレームワーク/ミドルウェア | OS など |
|-------------|-------------|-------------------------------|--------|
| 可視化アプリケーション | Swift 5.1.3 | UIKit, SwiftUI | iOS |
| 可視化アプリケーション | JavaScript | ReactNative, JavaScriptCore | iOS |
| GIS 化データ管理 | pSQL | PostgreSQL 12.1 + PostGIS 3.0 | CoreOS |

6.6 Docker

Docker は OS 仮想化システムの一つである。VirtualBox や VMWare Workstation などの完全仮想化ソフトウェアとは異なり、ゲスト OS の命令セットをホスト OS のカーネルの命令セットにコンバートすることにより仮想化を実現している。これにより、ハードウェアの仮想化を伴わないためオーバーヘッドが少なく機械学習のような計算量の多い課題に適していると言える。本研究では、機械学習システムの実行環境として採用をした。—— 主に、Python の実行環境系やライブラリなどを Ubuntu18.04 ベースのコンテナを作成した。このコンテナには本研究で用いた機械学習システムが依存する TensorFlow の実行環境が用意されている。

6.7 本研究で用いる環境定義アルゴリズム

ここにアルゴリズムを示す

```
for Routelist
    評価関数
end for
```

6.8 環境定義クラス

プログラム 6.1: 環境を構築するクラス

```

1
2 class CATPEnvironment:
3
4     def __init__(self, evaluator, step_t):
5         self._evaluator = evaluator
6         self._step = step_t
7         self._rewards = [ ]
8
9         # Register or memory agent action flow.
10    def act(self, act_obj):
11        res = act_obj.run()
12        rewards = self._evaluator(res)
13
14        self._rewards.append(rewards)
15
16        # Calculate environment diff and save environment.
17    def commit(self):
18        return sum(self._rewards)
19
20        # Clear all environment to initial.
21    def reset(self):
22        total_reward = sum(self._rewards)
23
24    def _capacity_validate(self, route_id):
25        if True:
26            return -10

```

プログラム 6.2: 行動を定義するクラス

```

1
2 class CATPAction:
3
4     def __init__(self):
5         pass
6
7     def reset(self):
8         print("Reset environment")
9
10        # Register agent action to environment
11    def act(self, act_obj):

```

```

12     pass
13
14     # Calculate environment diff
15     def commit(self):
16         print("Committing environment...")
17
18     def _capacity_validate(self, route_id):
19         if True:
20             return -10

```

プログラム 6.3: DQN の深層学習部分を形成するモデル

```

1 from tensorflow.keras.models import Sequential
2 from tensorflow.keras.layers import (
3     Dense,
4     Activation,
5     Flatten,
6     Convolution2D,
7     Permute
8 )
9 from tensorflow.keras.optimizers import Adam
10 import tensorflow.keras.backend as K
11
12
13 def dqnmodel():
14     model = Sequential()
15
16     # (width, height, channels)
17     model.add(Permute((2, 3, 1), input_shape=input_shape))
18
19     model.add(Convolution2D(32, (8, 8), strides=(4, 4)))
20     model.add(Activation('relu'))
21     model.add(Convolution2D(64, (4, 4), strides=(2, 2)))
22     model.add(Activation('relu'))
23     model.add(Convolution2D(64, (3, 3), strides=(1, 1)))
24     model.add(Activation('relu'))
25     model.add(Flatten())
26     model.add(Dense(512))
27     model.add(Activation('relu'))
28     model.add(Dense(nb_actions))
29     model.add(Activation('linear'))
30
31     return model

```

6.9 本実験における前提

本実験では道路交通網をソフトウェアプログラムにより擬似的に再現する。その際に、実験を短略化させ結果が発生する要因を掴みやすくするため以下の条件を適用する。

- 全ての道路は上下線双方向の通行が可能とする。
- 通行止めの発生は考えないものとする。

表 6.2: 本実験で適用する条件

| 条件 | 理由 |
|---|------------------------------------|
| 全ての道路は上下線通行可能 通行止めは考えない 本研究の趣旨と合わないため | 片道通行は複雑性が増すため 事故や災害などのアクシデント対応は |

- 信号による制御などは考えないものとする.

第7章 評価

本章では，提案システムの評価について述べる．

7.1 評価内容

本研究の実験において以下の項目について評価を行う．

- 人間の状態すなわち目的をレベル別に何個満たせたか
- 1km メッシュにおける人口の変異を時系列方向に観測し，全体的に移動が起きているか

第8章 結論

本章では、本研究のまとめと今後の課題を示す。

8.1 本研究のまとめ

この章では本研究のまとめについて述べる。

8.2 本研究の課題

本研究では、自動車に限りモビリティの利用者の満足度を高める経路制御を行う機械的な手法に取り組んだ。しかし、

8.2.1 想定環境の貧弱さ

本研究では、事前に選び出した幹線道路とその周辺にある主たる施設のみの仮想環境をソフトウェアで再現した。しかし、実際には交通機関は自動車意外にもあり鉄道やバスなどの大量輸送型の交通機関

8.2.2 学習器の連携

本研究では深層強化学習が利用者の満足度を高めるという視点で制御を行えるか実験を行った。ただ、実際に本研究の手法を適用する場合、自動車単体で行動は現実的に不可能である。従って、DQNによる強化学習器であるエージェントを複数設置する必要がある。下の図に示すように環境 ε を複数のエージェント α , β , Σ , Δ で共有するマルチエージェントモデルにおいて全体最適が取れるかを検証する必要がある。なお、環境 ε はそれぞれのエージェントから行動 a を受け取り、報酬 r を返す。

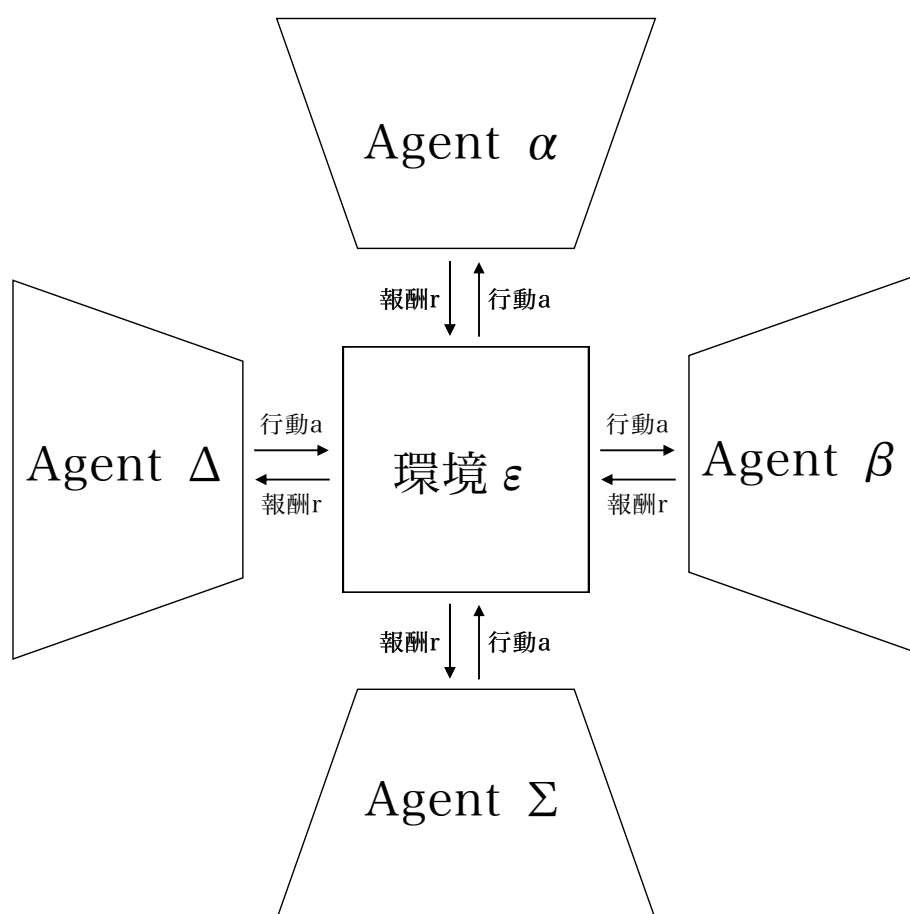


図 8.1: 複数エージェントで環境を共有する概念図

付 録 A APPENDIX

A.1 本論文で用いる機械学習関連の略語

- ANN - Artificial Neural Network
- DNN - Deep Neural Network
- CNN - Convolutional Neural Network
- DQN - Deep Q Network
- MaaS - Mobility as a Service

謝辞

本論文の執筆にあたり、ご指導頂いた慶應義塾大学環境情報学部村井純博士、同学部教授中村修博士、同学部教授楠本博之博士、同学部准教授高汐一紀博士、同学部教授三次仁博士、同学部准教授植原啓介博士、同学部准教授中澤仁博士、同学部準教授 Rodney D. Van Meter III 博士、同学部教授武田圭史博士、同大学政策・メディア研究科特任准教授鈴木茂哉博士、同大学政策・メディア研究科特任准教授佐藤 雅明博士、同大学 SFC 研究所上席所員齊藤賢爾博士に感謝致します。

特に齊藤氏には重ねて感謝致します。研究活動の中で様々な視点での助言をいただきました。

政策・メディア研究科 阿部 涼介氏に重ねて感謝いたします。私が同合同研究会に所属してから今日まで、研究活動を支援し

続けて、村井研究室 NECO の松本 光生氏、風間 宏治氏、宮元 眺氏は長く NECO としての生活を共にし、日々の研究生生活の助けとなりました。彼らと過ごした日々は私の興味範囲を広げると共に多くの刺激を得ることができました。深く感謝いたします。また、梶原留衣氏、渡辺 聡紀氏、木内 啓介氏、後藤 悠太氏、倉重 健氏、九鬼 嘉隆氏、内田 溪太氏、山本 哲平氏、吉開 拓人氏、金城 奈菜海氏、長田 琉羽里氏、前田 大輔氏には NECO のグループリーダーである私を様々な面から支えて頂き心から感謝しています。彼らの協力なしには研究グループの運営はできませんでした。

学外で私を様々な面で支えてくださった佐野 美月氏に深く感謝申し上げます。

私に、この卒論を書く学びの場を提供しこの世に生を受けてから今日まで私を支えてくださった。両親並びに弟に深く感謝いたします。

最後に、私をここまで成長させる基盤となった日本国並びに全地球に感謝いたします。

参考文献

- [1] Lacy Peter and Rutqvist Jakob. サーキュラー・エコノミーで競争優位性を確立する. https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/ja-jp/PDF_4/Accenture-Waste-Wealth-Exec-Sum-JP.pdf, 2019.
- [2] Silver David, Schrittwieser Julian, Simonyan Karen, Antonoglou Ioannis, Huang Aja, Guez Arthur, Hubert Thomas, Baker Lucas, Lai Matthew, Bolton Adrian, Chen Yutian, Lillicrap Timothy, Hui Fan, Sifre Laurent, Driessche George van den, Graepel Thore, and Hassabis Demis. Mastering the game of go without human knowledge. <https://www.nature.com/articles/nature24270.epdf>, 2017.
- [3] Shibata Katsunari. Learning of exploration behavior by reinforcement learning. <http://shws.cc.oita-u.ac.jp/~shibata/pub/SSI05.pdf>, 2008.
- [4] F. ROSENBLATT. The perceptron: A probabilistic model for information storage and organization in the brain1. https://pdfs.semanticscholar.org/5d11/aad09f65431b5d3cb1d85328743c9e53ba96.pdf?_ga=2.74309584.311052074.1580025098-406780135.1580025098, 1958.
- [5] Mnih Volodymyr, Kavukcuoglu Koray, Silver David, Rusu Andrei A., Veness Joel, Bellemare Marc G., Graves Alex, Riedmiller Martin, Fidjel Andreas K., Ostrovski Georg, Petersen Stig, Beattie Charles, Sadik Amir, Antonoglou Ioannis, King Helen, Kumaran Dharshan, Wierstra Daan, Legg Shane, and Hassabis Demis. Human-level control through deep reinforcement learning. <https://www.nature.com/articles/nature14236>, 2015.