

## 強化学習を用いた自立型掃除ロボットの進路決定に関する研究

日大生産工(院) ○川本 晃平  
日大生産工 豊谷 純, 大前 佑斗

### 1. まえがき

近年、様々な業界において自動化が急速に進んでおり、自動運転車が代表的である。ここで、掃除に対する自動化を考えると以前からロボット掃除機というものが存在していたが、自家用での床清掃などかなり限定的な条件であった。しかし、LiDAR(Light Detection and Ranging)と呼ばれる自動運転技術に欠かせないセンサーが自動運転車の進展とともに低価格化<sup>(1)</sup>が進んでいる。そのため、大規模商業施設などでLiDARを装着した自立型の掃除ロボットが使用されることが想定される。広範囲を複数の掃除ロボットが清掃する際、既存の進路決定アルゴリズムでは非効率なものになってしまう。そこで本研究では、複数の掃除ロボットが広範囲を清掃する際の効率的な清掃をマルチエージェント強化学習を用いて検討をする。

### 2. シミュレーション技術概要

#### 2.1 強化学習環境

本研究では、強化学習を行う際の環境としてOpenAIが公開しているOpenAI Gymを用いて環境の作成を行う。Gymは、強化学習用のプラットフォームであり、様々な環境が公開されており、強化学習アルゴリズムのテストに用いられる。

#### 2.2 MADDPG アルゴリズム<sup>(2)</sup>

本研究では、Multi-Agent Deep Deterministic Policy Gradientと呼ばれる強化学習アルゴリズムを用いる。MADDPGは、トレーニング時には全てのエージェントから情報を取得し学習し、学習されたモデルを使う際はエージェントそれぞれが取得できる情報を用いて行動する。そのため、複数エージェントの学習に向いている。

### 3. シミュレーション方法

掃除ロボットの進路決定を強化学習するため、Gymを用いてシミュレーション環境の作

成を行う。環境作成の際、留意する点として清掃する際、汚れやすい場所、汚れにくい場所が分かれることから図1の青から黄に移るにつれ堆積したゴミの量を変化させる頻度マップを用いる、あるいは図2の小さい黒点はゴミを示し、学習を行うごとにゴミ配置場所を確率ごとに変えることで様々な状況下において効率的な清掃を可能とすることを目的とする。また、エージェントが得られる情報として、図3では、青色の線がLiDARの検知を行う範囲を示しており、ゴミまでの距離、頻度マップを使った際は距離と量の情報を取得する。これらの条件から複数の環境作成を行う。これらの環境を作成後、経路決定を学習するためのアルゴリズムとしてMADDPGを実装する。比較として、単エージェントの学習に用いるDDPGまた、既存の経路決定のアルゴリズムとしてスパイラル走行が基本となっており、これらと比較検証を行う。

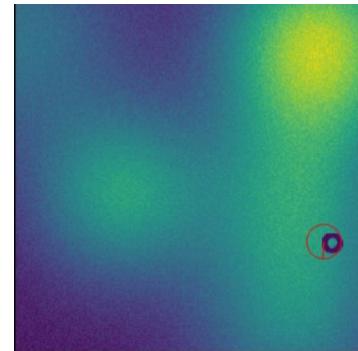


図1 ゴミの頻度マップ例<sup>(3)</sup>

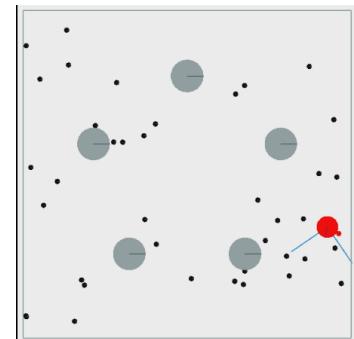


図2 ゴミの配置例<sup>(3)</sup>

Research on the course decision of a self-supporting cleaning robot using reinforcement learning

Kohei KAWAMOTO, Jun TOYOTANI and Yuto OMAE

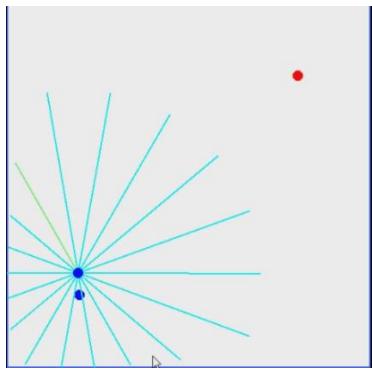


図3 環境上の LIADR 例<sup>(4)</sup>

#### 4. シミュレーションによる検討事項

環境を作成する際の検討事項として、制限時間や報酬を変更することでどのような経路に変わるかを検討する。図4、図5では青点がエージェントとして掃除ロボット、赤点がゴミを示している。図4では、二つのエージェントがあり、ゴミが左方に偏って点在している。この場合、ある制限時間内にゴミを取得した数を報酬とするとその制限時間が極端に短い場合、右側のエージェントは近くにあるゴミを取得しようとする。また、制限時間がある程度延ばすと右側のエージェントは左側のゴミを取得しに行くはずである。また、さらに制限時間を延ばすと右側のゴミから取得していくと考えられる。このような状況下では、制限時間が結果にどのような影響を及ぼすか確認をする必要がある。

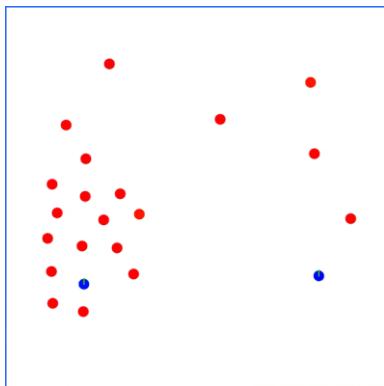


図4 二つのエージェントとゴミの配置

図5では、三つのエージェントがあるが、この場合、制限時間に関わらず左側と上側のエージェントが左側のゴミを取得し、右側のエージェントは右側のゴミを取得すると考えられる。この三つまたはそれ以上のエージェ

ントを利用する際、特に協調の動きが見れると考えられ、このような状況の下では単エージェントに用いる強化学習や既存の経路決定アルゴリズムと比較して掃除効率に大きな差が生まれるのではないかと考えられる。

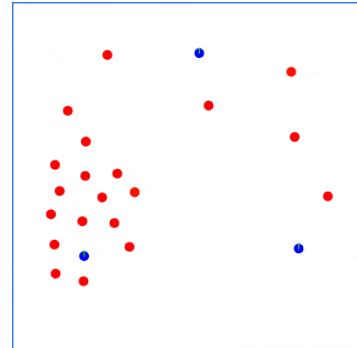


図5 三つのエージェントとゴミの配置

また、大規模商業施設などの広範囲を模すため、グリッドのサイズ、ゴミの量、エージェントの数を多くしていった際に動きに変化が生じるのか、障害物を置いた際どのような変化が生じるのかについて検討を行う。

#### 5. まとめ

本研究では、複数台の掃除ロボットの経路決定を強化学習により学習を行うため、ロボットの台数が何台であるか、ゴミの分布はどのようにあるかなどの条件により様々な動きになると想定される。また、ゴミを单一にせず、ゴミにより報酬値を変更する、通過点を設定する、スタート地点、ゴール地点の設定などによりどの程度意図的に操作できるかも実運用する際には重要になると考えられる。

#### 参考文献

- 1) 「Iris」 LUMINAR  
<https://www.luminartech.com/products/>  
 (最終閲覧日: 2020/10/09)
- 2) Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, Igor Mordatch (2017) Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments
- 3) 岡 瑞起・池上 高志・ドミニク チェン・青木 竜太・丸山 典宏(2018)『作って動かすALife—実装を通した人工生命モデル理論入門』オライリージャパン
- 4) Yudai Sadakuni 「OpenAI Gym original environments」 GitHub  
<https://github.com/Sadaku1993/gym-pathplan> (最終閲覧日: 2020/10/06)