ISSN 2186-5647

5-29

マルチエージェント強化学習を用いた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

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[1]が進んでおり、また人件費の高騰から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OpneAI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。また、Gymを用いて作成された既存の環境であるMulti-Agent Particle Environmentを利用することで環境作成に係る時間を軽減する。

2.2 MADDPGアルゴリズム[2]

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。

2.3 DDPGアルゴリズム

本研究では、MADDPGの比較としてDDPG(Deep Deterministic Policy Gradient)を用いる。連続値制御問題に対して有効な手法である。従来の方策勾配法では連続値を扱う場合に確率分布が必要であったが、DDPGによりある状態に対して一意のアクションを求められるようになっている。

3. シミュレーション方法

本研究のシミュレーション環境の作成において、想定する状況は以下のものである。前提として、既存の家庭用の掃除ロボットを複数台使用する。掃除ロボットは、既に掃除エリアをマッピング済みである。マッピングはグリッド状に別々に行われており、縦横10×10の合計100グリッドを使用する。掃除ロボットは上下左右の4方向にのみ移動可能である。それぞれのグリッドにはゴミの量を割り振る。また、エージェントが得られる情報は各グリッドにおけるゴミの量と座標、各エージェントの座標である。

上記の想定する状況をもとにGymを用いてシミュレーション環境の作成を行う。それぞれのグリッドに割り振るゴミの量は図1のヒートマップを使用する。図1は濃淡で、ゴミの量を示しており濃い方が多く、薄い方が少なさを示している。ヒートマップの複雑さが学習の難度に影響するため、随時調整を行う。図2は図1のヒートマップを利用した時のテスト環境である。黒丸がエージェントの位置を示しており、青、緑、赤の順にゴミの量が多くなっている。

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と比較検証を行う。

Research on the course decision of a cleaning robot using multi-agent reinforcement learning

Kohei KAWAMOTO and Jun TOYOTANI

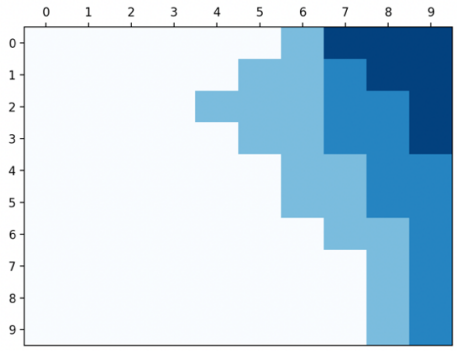

図１ ゴミの頻度マップ例

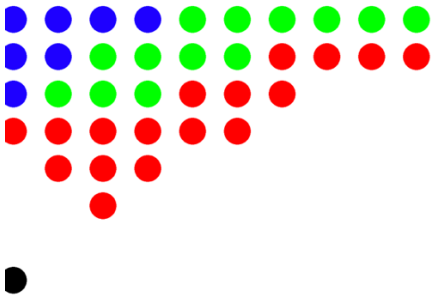

図２ ゴミの配置例

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、頻度マップの複雑さと報酬、制限時間が挙げられる。頻度マップは、座標と量の変数があり、マップは固定しているため変更し得る変数は量となる。図1、図2ではゴミの量を３段階で示しているが、精度の点で懸念が残る。しかし、より細かい値にする場合学習が難しくなることは強化学習の性質から明白である。このトレードオフの関係を考慮して環境の作成を行う必要がある。報酬の設定に関して、現状では1ステップごとにマイナスの報酬を付け、ゴミを取得した際に量に応じてプラスの報酬を与えることを考えている。どの程度プラスあるいはマイナスをするかを調整する必要がある。また一度のシミュレーションにつき、終了条件としてステップ数を調整する必要がある。

また、マルチエージェント強化学習を単エージェントに対する強化学習との違いとして、協調学習がある。通常の強化学習では、エージェントがそれぞれ自身の報酬を最大化する行動を行うため、全体での報酬の最大化にならないという問題があった。これを解消したものがマルチエージェント強化学習であり、互いのエージェントが協力し、全体の報酬の最大化を行う。しかし、作成する環境によっては通常の強化学習とマルチエージェント強化学習との差に有意な差がない可能性がある。ここで、検討したいのは図３のような状況である。上側、右側にゴミがまとまってある環境である。２つのエージェントが左下にあると仮定した場合、ステップ数の制限にもよるが通常の強化学習を行うと上側のゴミを互いに奪い合うようにして全体の報酬の最大化ができない可能性がある。このような環境下での違いを検討する必要がある。

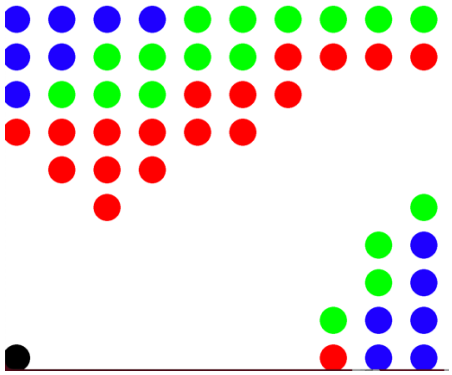

図３　検討するゴミの配置

そして、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も検討を行う。

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作成した環境の元でマルチエージェント強化学習を行なった場合、強化学習では見られない協調行動を行うと考えられる。現状、マルチエージェント強化学習は研究段階であり、産業での使用は検討段階である。しかし、協調行動は最適化を行う際に重要な要素となると考えるのでシミュレーションに基づいて検討を行い有意性の確認をしたい。

参考文献

1) 「Iris」LUMINAR https://www.luminartech.com/products/ (最終閲覧日：2021/10/12)
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