5X-02

移動ロボットの行動モデル獲得のため多様なシミュレーション環境構築と行動学習を自動で行うシステムの開発

海保　諒†　　鶴田　龍登‡　　森岡　一幸‡
明治大†　　明治大‡　　明治大‡

1.　緒言

深層強化学習に基づく行動モデルを用いて走行する移動ロボットは状況に応じて目的地までの適切な行動を出力できるため、事前に環境地図が必要なく未知の環境での走行が期待できる。行動モデルの訓練は一般的にシミュレータにて行うため実環境を模した環境が必要になるが、手作業による環境構築では、環境地図を用いずに走行できる走行システムの利点が失われるため、自動で構築する必要があった。そして、人が眼で進むべき道を判断するように行動を学習させるべく、三次元環境においてカメラ画像を用いた行動学習を行う必要もあった。よって、本研究の目的は環境の構築から行動の学習まで一貫して行うシステムの開発である。そして、訓練した行動モデルを用いて自律走行に使用可能であるか検証を行う。

2.　環境構築と行動学習を行うシステム

図１は本研究にて開発を目指すシステムの全体図である。著者らの従来研究において、WFC によって多様化させたシミュレーション環境での行動学習を行い、生成した行動モデルでのシミュレーションに成功している[1]。本研究においても同様のシステムを改良し用いることとする。

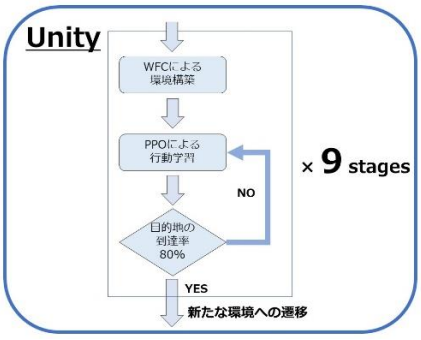

図１　システム全体図

Development of a learning system of action models for mobile robots based on automatic building of diverse simulation environments.
† Ryo KAIHO Meiji University
‡ Ryuto TSURUTA Meiji University
‡ Kazuyuki MORIOKA Meiji University

3.　シミュレーション環境構築

3.1　WFC について

WFC とは入力した bitmap に局所的に類似した bitmap を生成するアルゴリズムである。本来は二次元画像に対して適用するアルゴリズムであるが、Tessera[2]と呼ばれるツールを用いることで、Unity 上の三次元環境に対しても適用が可能となった。本研究では、この Tessera を用いた Unity 上にシステムを開発していく。

3.2　WFC による環境構築

従来研究では単なる壁のみ存在する環境を構築した。本研究においても、同様の手法にてシミュレーション環境の構築を行うが、より多様な環境を構築すべく図２に示すタイルを用いる。一部のタイルは、茨城県つくば市で毎年開催される「つくばチャレンジ」のシミュレーション環境として開発された VTC on Unity[3]の素材を一部使用させていただいた。これらタイルに定める接続ルールは道（今回のタイルでは薄い灰色の箇所）同士の接続は可能といったものである。このルールを定めることで、タイルが適切に接続され、シミュレーション環境に道が構築される。

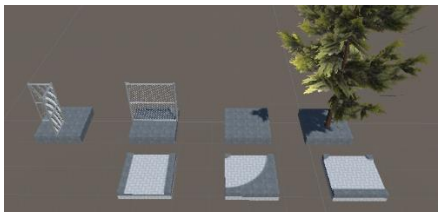

図２　本研究で用いるタイル

4.　Unity による行動学習

本研究で用いる移動ロボットは単眼カメラを搭載した前後左右の移動が可能である車輪型移動ロボットを想定している。

4.1　強化学習システム

状態を入力として、行動に対応した離散値を出力する行動モデルにより走行する。状態は単眼カメラより得られる二次元画像（図３）や移動ロボットの速度、そして移動ロボットと目的地それぞれ

のX座標（横）とZ座標（奥行き）から計算した距離の合計４つの情報を入力とする。図３は移動ロボットの視点である。この位置に搭載した単眼カメラより二次元画像の取得を行う。行動は前後左右への移動と移動しないという５つから構成し、行動モデルはこれら移動に対する離散値を出力するものとする。

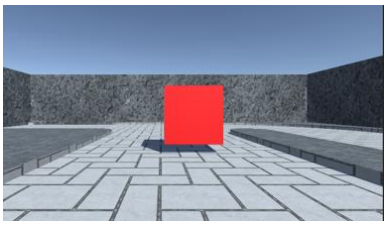

図３　移動ロボットの視点画像

4.2　行動モデルの訓練方法

前述した行動モデルはカメラ画像の処理を含むためCNNと全結合型のニューラルネットワークを組み合わせて構成される。シミュレータ内で移動ロボットを走行させ、ニューラルネットワークの訓練をすることで、行動モデルを獲得する。シミュレータとして用いるUnity上にて、目的地に到達したら+1.0、障害物に衝突したら-1.0の報酬を与える。また、0.02秒ごとに目的地に近づく行動をした場合と道上に移動ロボットが存在する場合は+0.01 の報酬を与える。一方で、遠ざかる行動をした場合と道上に移動ロボットが存在しない場合は-0.01 を報酬として与える。よって、目的地を目指しつつも道上から外れない行動を学習することが期待できる。

4.3　学習環境

図４にて示す９つのステージにてML-Agentsを用いた行動学習を行う。図１のシステム全体図で示したように目的地への到達率によって、新たに環境を構築することで学習を繰り返す。また、青い立方体が移動ロボットを表しており、赤い立方体が目的地である。

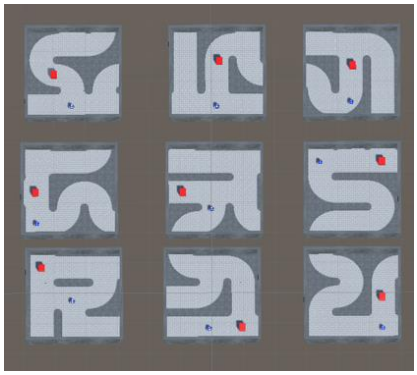

図４　学習環境

4.4　学習条件

図１に示したように行動学習を行う。また、学習の終了条件は全てのステージにて少なくとも一度は環境遷移が行われ、累積報酬が正となるタイミングとする。

4.5　結果

学習した行動モデルを用いて、図５に示す環境にてシミュレーションを行った。試行回数は１００回とし、目的地への到達確率を計算する。また、スタート地点とゴール地点は道上からランダムに選択した任意の点とする。結果は６８％であった。図６に行動学習時の累積報酬を示す。

図５　実験環境

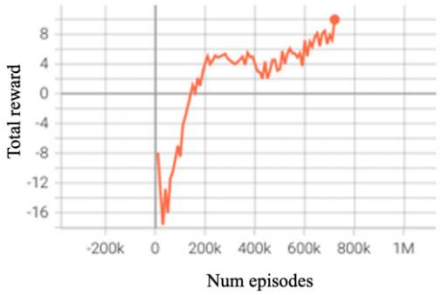

図６　累積報酬の推移

5.　まとめ

実験の結果、走行成功率は約７割となり従来研究と同水準であった。今後は様々なタイルを用いて実環境に近い多様な環境の自動構築をするとともに、適切な報酬設定や強化学習手法の改善を行うことで、成功確率９割を達成することを目指す。

参考文献

[1] 海保諒，鶴田龍登，森岡一幸，“シミュレーション環境の多様化による移動ロボットの汎用的な行動モデル獲得”，計測自動制御学会SI部門講演会 SI2021, 1H5-01, 2021.

[2] Adam Newgas, “Tessera: A Practical System for Extended WaveFunctionCollapse”, Proc. of the 16th International Conference on the Foundations of Digital Games (FDG'21), Article 56, pp.1-7, 2021.

[3] Ryodo Tanaka, vtc_unity, https://github.com/Field-Robotics-Japan/vtc_unity, 2020.