

ライントレースロボットを用いた強化学習の教材開発

日大生産工(院) ○諏訪 晃也 日大生産工 柳澤 一機

1 緒言

近年, 人工知能技術は様々な実問題に適用されつつあり, 人工知能技術とロボティクスの融合が進んでいる¹⁾. そのため, それらを取り扱える技術者の育成が急務となっている.

現在, 人工知能や機械学習を学ぶための教材は, 実環境での学習の困難さからシミュレーションのみで完結するものが多く, 実機での制御までは考慮されていない. しかし, Julian Ibarzらはシミュレーション上で学習した内容は, リアリティギャップによってうまく機能しないことを指摘している²⁾. ここで言うリアリティギャップとは, シミュレーションでは表現しにくい電池残量によるモータのトルクの変化や, 制御部分の計算遅延といったシミュレーションと現実での差異である. したがって, 実環境で学習できる教材が非常に重要である.

そこで, 本研究では数ある機械学習の中から, ロボットの動作生成と相性の良い強化学習を題材とし, 状態の観測が容易であるライントレースロボットを対象にした教材を開発する.

2 強化学習とは

強化学習とは, 良い状態と悪い状態のみを決めておくことで, その過程を自動的に学習し, 良い状態になるまでの動作を獲得する機械学習の一種である.

強化学習では, エージェント(動作を学習する対象)に対して「環境」(動作を行う空間)から「状態 s_t 」(エージェントが環境から得られる情報)を得る. エージェントは「状態 s_t 」において, あらかじめ設定された「行動 a_t 」(エージェントが環境内でとる行動)の中から決められたルールに従って行動を決定する. 「行動 a_t 」の後の「状態 s_{t+1} 」を判定し, その状態の良し悪しによって, 「状態 s_t 」において「行動 a_t 」を選択する価値である最適行動価値関数 $Q(s_t, a_t)$ に対して「報酬 r 」が加減算される仕組みである.

本研究では, 強化学習のアルゴリズムとして最も簡単であるQ学習という手法を取り扱う.

Q学習では, 基本的に最も $Q(s_t, a_t)$ の値が高い「行動 a_t 」を選択する.

3 教材概要

本教材は, 強化学習を利用してライントレースロボットを制御することで, 強化学習について学び, 経験を通じて実環境で機械学習を用いてロボットを制御するために必要な要素についての知識を得ることを目標とする.

強化学習は, その性質上試行錯誤を繰り返すシステムとなっているため, その全てを実環境で行う事は困難であり, 時間もかかる. そのため, まず初めにシミュレーション上でロボットに学習させ, 学習した内容を実環境に適用する.

本教材では, シミュレーション, 実環境によって構成され, それぞれの内容を教科書を通じて教示する方法をとる.

第1段階として, シミュレーションを動かしながら強化学習の概要について学び, 強化学習の重要な要素である行動, 状態, 報酬について体験を通じて理解してもらう.

第2段階として, シミュレーションで学習した内容を実機に適用し, シミュレーションと実機での強化学習の取り扱いの違いや, リアリティギャップによる影響など, 体験を通じて理解させ, それらの解決方法について考察するという内容となっている.

4 教材内容

4.1 ライントレースシミュレーション

本教材のシミュレーションの概要を説明する. シミュレーションにはpythonを使用しopen-cvによって描画される. 描画される内容を図1に示す.

シミュレーション上でのライントレースロボットは, センサ位置の明度を8bit(1024段階)で読み取り, 常に4px/行動の速さで前進する.

初期設定では状態として35度刻みで設置された3つのセンサで読み取った明度がそれぞれ黒か白かを判別する. 行動は右に10度旋回しながら前進と左に10度旋回しながら前進の2つから選択する. そして報酬は, 真ん中のセンサが黒を読み取っている時に正の報酬を与える.

これらのセンサの個数や位置, 状態・行動・報酬の3要素は, センサで読み取る状態に灰色

Development of teaching materials for reinforcement learning using line trace robots

Koya SUWA, Kazuki YANAGISAWA

を加える, 行動に巡回せずに前進する行動を加えるなど自由に変更する事ができる. 定めた状態・行動・報酬の条件において実際にロボットが学習する様子を観察することで, 経験を通じて強化学習の重要な3要素について理解を深める内容となっている.

シミュレーション終了時にはロボットが学習した $Q(s_t, a)$ が保存され, 学習結果の確認や, 実機への適用が可能である.

すでに実際にシミュレーションと教科書を用いて授業を行い, 強化学習の理解が深まることを確認している³⁾.

4.2 実環境でのライントレース

開発したライントレースロボットの概要を説明する(図2).

ロボットの制御にはraspberrypi3+Bを使用しており, シミュレーション同様pythonによってプログラミング可能であるため, シミュレーションと同じ感覚で状態・行動・報酬を設定でき, シミュレーションと実環境でのQ学習の挙動の違いを体験しながらロボットを制御できるような仕組みになっている.

基本的にシミュレーションで学習した $Q(s_t, a)$ を使用して学習させる.

センサ部は5度刻みに自由にセンサを取り付けることが可能であり, 位置と個数を変更することが可能である. それにより, シミュレーションと同じセンサ数, 位置を再現できる.

駆動部にはハイスピードサーボを使用することで, シミュレーションと同様の角度制御で動かすことが可能である. タイヤは摩擦の大きいシリコンと摩擦の小さいゴムの2種類から選択することができ, タイヤが滑ってしまう事によって強化学習にどのような影響を及ぼすかといったリアリティギャップの要因にもなる「床との摩擦」を体験する事ができる⁴⁾.

5 結言

機械学習の一つであり, ロボットと相性の良い強化学習を題材とし, 状態の観測が容易であるライントレースロボットを対象にした教材の開発を行った. シミュレーションを通じて, 強化学習の重要な要素である行動, 状態, 報酬の3要素について体験を通じて理解する. その後, シミュレーション上でロボットに学習させた内容を基に実環境でも学習させることで, リアリティギャップについても体感できる教材を開発した.

今後は, 実際に作成した教材を使って授業を行い, その効果について検証する.

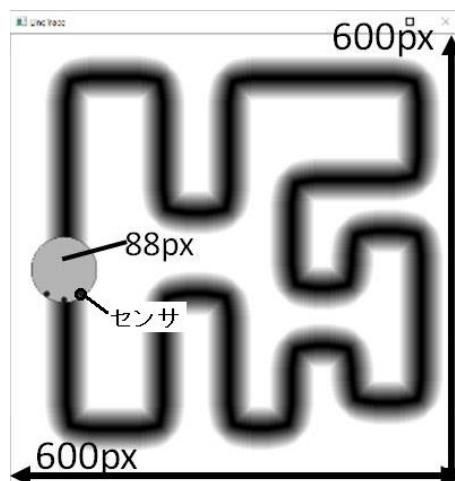


Fig.1 ライントレースシミュレーション⁴⁾

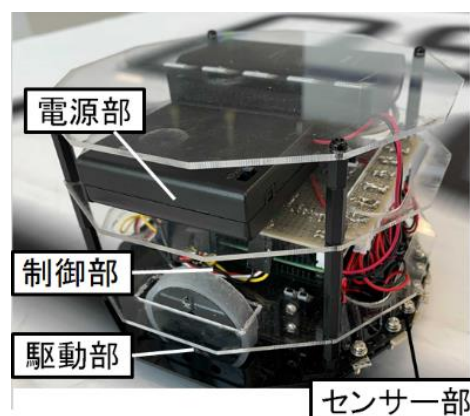


Fig.2 開発したライントレースロボット⁴⁾

参考文献

- 1) 比戸 将平, 人工知能技術のロボット産業応用, 日本ロボット学会誌 Vol35, No.3, (2017)pp.186 - 190.
- 2) Julian Ibarz, Jie Tan, Chelsea Finn, and more, How to train your robot with deep reinforcement learning, lessons we have learned, The International Journal of Robotics Research, Vol. 40, No. 5, (2021)pp.698-721.
- 3) 諏訪 晃也, 柳澤 一機, ライントレースロボットを用いた強化学習の学習教材の開発, 情報処理学会, 情報処理学会第83回全国大会(2021)4ZF-03.
- 4) 諏訪 晃也, 柳澤 一機, ライントレースロボットを用いた強化学習の教材開発, ヒューマンインターフェース学会, HumanInterface2021SENDAI(2021)PD4-1.