

ロボメックの指針

Shibuya Atsushi 渋谷享史

2024 年 3 月 3 日

要約：本研究は、従来の制御方法では不確実性が含まれ対処しきれない問題を、強化学習を用いて得た制御方法に部分的に置き換えることで、全体の設計を維持しつつ、そのロボットの制御が抱えている問題に対処する手法を提案する研究である。

1 問題（諸言の部分の内容）

ロボットが移動・走行するに当たって、想定していない事象、例えばタイヤのスリップなどが発生することが起こりうる。ある程度は想定して、それを踏まえて設計することで対処がなされてきていたが、複雑な要因が絡んだ上で起きるため対処が難しいものなどがあり、またその対処には、設計の工夫を考える必要があるなど労力がかかり、時間的なコストもかかる。

それを克服する手段の一つとして、強化学習を適用する考え方がある。強化学習により、数々の経験から最適な行動を行うことで制御を得ることで、起こりうる様々な考慮しきれない外部要因による影響をも含んだ上での制御を得ることが可能となり、対応しきれなかった幅広い事象に対応可能な制御手法を獲得する研究が活発に行われている。具体的には、モデルの設計が複雑で困難な 2 足歩行ロボットの制御方策獲得に関する研究 [?] や、深層強化学習によって様々な状況に対応可能な制御を獲得した研究 [?], 深層学習によって、まるで意味を理解したような、柔軟で人が設計するには困難な動きをロボット自らが獲得した研究 [?] 等がある。

しかし、強化学習を用いてそのロボットの制御全体を置き換えることは容易ではなく、従来の制御全体を超える性能を得るためには、学習アルゴリズムのパラメータの試行錯誤などが必要となってくる。また、強化学習によって得られる制御は条件反射的な行動が多く、順序だてて複雑な動きを行うような制御を得るのは難しいという特徴 [?] があった。

そこで、既にある従来の制御手法をもとに、その機能ごとに、強化学習によって得た結果を用いる手法を提案する。この手法のメリットは、すでにある制御対象の動作の柔軟性向上を効率的に行えることにあり、従来の制御手法の一部分を取り替えるため、プログラムを最初から準備する手間が省略できる。そして、強化学習の適用においても、様々な環境に一度に対応するような方策を得ることは難しいが、機能単体ごとに学習させるため学習の行い方を決めやすく、複雑な動作を学習によって得るために労力をかけて学習方法を設計する必要もなく、学習を行いやすい。そして強化学習の特徴でありメリットである「経験から最適な結果につながるような行動」を得ることができ

るため、従来の手法で解決が難しかった特徴を、経験により考慮することで解決することが可能になると考えた。本研究では、その手法を検証する題材として、単純化された環境を走行する、協議会が行われている2輪移動ロボット「マイクロマウス」での左右旋回動作を選び、実験を行う。また、従来の制御手法としては、公開されているサンプルプログラムをもとに、迷路内をぶつからずに前進し続ける動作を追加したものを考える。各マスの移動を行うプログラムを基本に、取り付けられた4つのセンサから壁の状況を判断し、前進、左旋回、右旋回の行動を選び、各マスごとに移動しながらその行動選択を繰り返すことで前進し続けるというものである。マスの大きさはすべて決められた値であり等間隔で並んでいるため、パラメータの調整によって、マス一つ分の移動になるように設計されている。また、外乱などによるずれの補正を行うアルゴリズムを追加することで、壁にぶつからずに前進し続ける制御を行うものである。しかしこの制御ではタイヤのスリップなどによる誤差が生じてしまい、ある程度の誤差を含む想定通りのマス移動を行うことが前提の従来の制御プログラムでは、誤差がたまり壁にぶつかってしまう問題があった。また、直進区間がある場合は、左右のセンサの値からマスの中心に戻るように補正をかけることができていたが、それが行える区間のない旋回動作が連続するような箇所では誤差の修正ができず、対処が考えにくい問題が見られていた。本研究では、その問題を提案手法で改善することを目的とし、本手法の有用性を評価する。したがって本研究は、従来の制御方法では対処の難しい問題を、強化学習を用いて得た制御方法に部分的に置き換えることで、全体の設計を維持しつつロボットの制御が抱える問題に対処する手法を提案する研究である、

2 提案手法

本研究では、ロボットの各動作と同じ役割をもつ制御を強化学習によって獲得し、その制御に置き換えることで従来制御の対処しきれない問題を解決することを目指す。本研究で扱うマイクロマウスの従来制御では、連続した左右旋回が続く区間で補正がしきれないという問題が見られていたため、まず、従来の制御手法でロボットを走行させ、そのロボットが到達できた距離を記録する。その後、強化学習によって得られた制御を用いて同じコースを走行させ、そのロボットが到達できた距離を記録する。その到達距離の比較によって、従来の制御の抱えていた問題の改善に有効であることを示すことで、提案手法の有用性を評価する。実験はシミュレーション上で行う。

2.1 従来の制御手法

従来の制御手法として扱うものについて説明する。基本的にマス目のサイズが定められている環境を走行するため、1マス前進、1マス右にすすむ、1マス左にすすむ、という3つの動作が基本的な動作である。具体的にそれらの動作は、ロボットから見た次のマスの壁の配置状態によって決まる。

- 前進：前方に壁がない場合、0.18[m] 前進
- 左旋回：左側に壁がない場合、0.09[m] 前進、90[度] 左旋回、0.09[m] 前進

- 右旋回：右側に壁がない場合，0.09[m] 前進，90[度] 右旋回，0.09[m] 前進
- 袋小路：前方，左右ともに壁がある場合，180 度旋回

となっている．基本的にはこれらの動作は，目標速度値を追従するような速度を出すことによって台形加減速を行うことで実現されている．しかし，この方法のみでは移動ごとに誤差を含んでしまう場合がある．そのためこれらに加え，以下のような位置補正処理が行われている．

- 補正 1（前進移動中）：左右のセンサの値の差から，マスの中心ラインと機体のズレを計算し，戻るよう補正．
- 補正 2（前進移動中）：前方のセンサの値から，壁に 0.09[mm] 以上近づいたらそれ以上は進まないようにする．
- 補正 3（前進移動後）：前方のセンサの値の差から，機体の向いている方向のズレを計算し，ズレが小さくなるようにその場で向きを整えることで補正．

この補正を加えることで，走行結果にある程度のズレが生じた場合も，それを補正することができるようになっている．しかし，このような補正を加えても補正しきれない場合があった．

2.2 強化学習による制御手法

次に，強化学習による制御手法について述べる．強化学習は，以下のような環境で学習を行った．壁がランダムに出現する環境で，左 90 度旋回動作と，右 90 度旋回動作を学習させた．状態空間は以下の表に示す．マイクロマウスに取り付けられたセンサの値と，車輪のモータに取り付けられたエンコーダからオドメトリによって推定される値を用いた．次元数は 12 である．（表を入れる）行動空間は，ロボットに搭載された左右の 2 つの車輪を回すモータへの制御入力とし，次元数は 2 である．これにより，考える壁の状態をすべて踏まえた上での，センサの値を用いて 90 度旋回を経験から行う方策の獲得を目指した．

（図を入れたりして説明をわかりやすくする）

2.3 実験方法

これらの制御によって，同じコースを走行させ，その到達距離を比較することで，提案手法の有用性を評価する．具体的には，図のようなコースを走行させる．また，この地面は基本的には摩擦係数が十分タイヤが滑らない値となるよう設定されているが，一定の確率でタイヤがスリップしてしまう摩擦係数に変化する．

このような設定によって，現実世界でのタイヤの滑りを再現した．この，一定確率でタイヤのスリップが生じてしまうコースで 2 種類の制御によってロボットを走行させ，スリップという外乱からの安定性を比較することで，提案手法の有用性を評価する．各制御を 100 回ずつ走行させ，到達距離の平均値を比較する．従来手法の制御である程度補正をかけられていたが，対応しきれなかった場合があるものに対して本手法を用いて置き換えた制御で改善できることを示し，提案手法の有

用性を示す.

3 結果・考察

同じフィールドを2種類の制御を用いて100回ずつ走行させた際の到達距離を表すグラフを以下に示す.

グラフより、(提案手法による制御の方が、ぶつからずに走行できている距離が長いことがわかる. よって、安定して走行できていることがわかる. このことから、従来の制御よりもスリップによる影響を受けにくいことがわかる.) また、その制御のパラメータを変更して実験を行った場合を以下に示す.

これにより、値が違っても本手法は有用であることがわかる.

4 結言

本研究では、従来の制御の問題を解決する一つのアプローチとして、強化学習によって得た制御を部分的に置き換える手法を提案した. マイクロマウス協議会のフィールドを使用し、従来の制御と、強化学習によって得られた制御に置き換えた制御を比較することで、提案手法である「一部を強化学習で置き換えることで従来の制御の改善を行う」の有用性を示した. 今後は、他の問題に対してもこの手法を適用し、その有用性を示すことができるかを検証する必要がある.

5 現状と今後の予定

PPOによる深層強化学習で、左右の旋回動作を学習させた.

台形加減速によって左手法を行わせるプログラムを従来の制御として用意し、その制御の左、右旋回時は学習結果を推論した値によって行動させるものを別に用意する. その2つで同じコースを走行させる. 連続旋回区間(ギザギザの地形)で、5%の確率でタイヤの摩擦を変化させるなどしてスリップを再現した環境を用意し、100回あたりのゴール率、壁接触率などを比較することで、この手法の有用性を主張する.

残っている作業

- 台形加減速での左右旋回のゲイン調整
- 学習結果を使って左右旋回するプログラム
- ランダムにスリップする連続旋回区間の環境の作成
- 比較実験
- 前刷りの仕上げ