2Q-03

強化学習を用いたコンピュータシミュレーション環境内でのロボットアームにおける運搬動作学習の基礎検討

山下　茜音[†1]　伊藤　悠大[†1]　景山　陽一[†1]　茨木　大基[†2]　廣瀬　聡[†3]

秋田大学[†1]　株式会社ネクストスケープ[†2]　日本ビジネスシステムズ株式会社[†3]

1. 背景・目的

近年，産業分野における人手不足を解消するため，産業用ロボットの導入が増加している[1]．中でもロボットアーム[2]は，作業の安全性および汎用性の高さから，精密な作業や複数のロボットによる協力作業など，人とロボットの分業に関して重要な役割を果たしている．

一方，農林水産省の報告によると，食品製造業では，過半数がロボット導入を検討したことがないと回答しており，他の業種と比較して導入が遅れている[3]．ロボット導入の課題として，ロボットを設置するスペースの確保が挙げられている．例えば，汎用品のロボットアームを人の代替として単に設置した場合，アームが壁にぶつかる，無理な姿勢で止まるなどの齟齬が生じる可能性がある[4]．そのため，実務環境を模したシミュレーション環境において，ロボットアームの動作を確認することは重要である．しかしながら，ロボットアームの動作を各々の実務環境および必要な動作に合わせて設定することは，専門性が高く必ずしも容易ではない．

そこで，実務に必要となる動作の学習を自動化し，シミュレートを行うことは，ロボットアームの導入検討において有用であると考える．本研究では，コンピュータシミュレーション環境内で，強化学習を用いて，明示的なプログラムなしにロボットアームが動作を学習する手法の開発を目的とする．本稿では，6 軸のロボットアームを作成し，強化学習を用いて対象物を目的位置まで運搬する動作について検討を行った．

2. 構築した環境およびオブジェクト

2.1. シミュレーション環境

Unity[5]を用いてコンピュータシミュレーション環境を構築し，Unity のプラグインである ML-Agents[6]を用いて強化学習を行った．

2.2. 作成したオブジェクト

A~I のパーツで構成される作成したロボットアームを図 1 に示す．各パーツの動作および制限角度を表 1 に示す．加えて，アームの運搬対象物(以下，ワークと表記する)，ワークの生成範囲(以下，始箱と表記する)および運搬の目的範囲(以下，終箱と表記する)を作成した．また，始箱および終箱(以下，箱と表記する)の各上部に

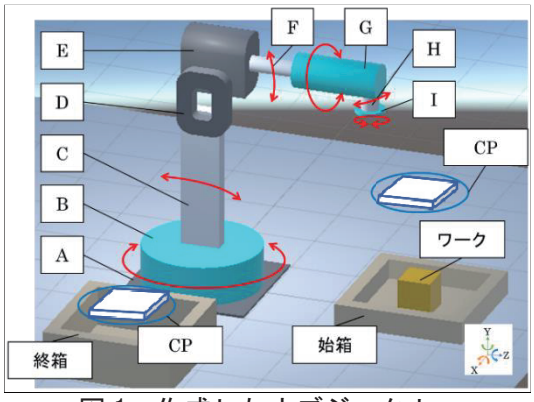

図 1. 作成したオブジェクト

表 1. 各パーツの動作および制限角度

名称	動作	制限角度
A	固定	
B	体を水平に回転させる	-100 度~100 度
C	体を前後に動かす	-30 度~50 度
D	動作なし	
E	腕を上下に動かす	-30 度~60 度
F	動作なし	
G	腕を回転させる	-30 度~30 度
H	腕の先を上下に動かす	-30 度~30 度
I	腕の先を回転させる	-180 度~180 度

衝突判定により報酬を与える範囲(以下，CP と表記する)を作成した．

3. 強化学習

3.1. 強化学習の概要および設定項目

強化学習は，動的環境と試行錯誤によりタスクを実行できるようにする手法であり，明示的なプログラムなしにタスクの報酬を最大化する一連の意思決定を行うことを可能にする[7]．

強化学習のアルゴリズムには，Proximal Policy Optimization(以下，PPO と表記する)を使用した．PPO は，実装が単純かつ一般的で高パフォーマンスであるという特徴をもつ[8]．学習における 1 episode は 800step とした．学習の終了条件は，ワークを終箱に入れた場合，ワークを床に落とした場合，または step 数が上限に達した場合である．step 数の上限は，5000 万 step とした．

3.2. 学習動作

ロボットアームがワークを始箱から終箱へ運搬する動作の強化学習を行う．ここで，ワークを終箱に入れた場合を成功条件とし，ワークを始箱または終箱以外に落とした場合，アームの先端(以下，先端と表記する)が箱に衝突した場合，ならびにワークが箱に衝突した場合を失敗条件とする．また，ワークを運搬する際は，①ワー

Basic Study of Learning Carrying Motion in Robot Arm Using Reinforcement Learning by Computer Simulation

Akane Yamashita[†1], Yudai Ito[†1], Yoichi Kageyama[†1], Hiroki Ibaraki[†2], Satoshi Hirose[†3]

†1 Akita University, †2 NEXTSCAPE Inc, †3 Japan Business Systems, Inc.

Copyright ©2024 Information Processing Society of Japan.
All Rights Reserved.

クへ近づく，②ワークを掴む，③ワークを離さず終箱に近づく，ならびに④ワークを離す，の 4 つの動作を必要とする．

3.3. 取得する観察値

観察値として A を除く B~I の各パーツ，ワークの座標(x, y, z)および角度の計 36 項目を取得した．加えて，先端から光線[9]を射出することで，光線に当たったオブジェクト(床，ワーク，始箱，終箱)と先端の距離および角度を取得した．光線を射出する様子を図 2 に示す．

4. 設定した報酬

4.1. 成功条件および失敗条件における報酬

運搬動作を学習させることを目的として，成功条件を満たした場合に最大の報酬 (+1.0)を与え，失敗条件を満たした場合に最低の報酬(-1.0)を与えた．

4.2. ワークを掴んだ場合および CP 通過の報酬

3.2 節で挙げた②の動作の実現を目的として，ワークを掴んだ場合に報酬(+0.5)を与えた．また，③の動作の実現を目的として，箱の各上部に CP を設置し，ワークが CP を通過した場合に報酬(+0.5)を与えた．始箱上部にある CP は，ワークを掴んだ後，アームを上に持ち上げる動作に報酬を与えるために設置した．また，終箱上部にある CP は，終箱に入れる際に先端が左右にぶれることなくアームを降ろす動作に報酬を与えるために設置した．なお，各 CP の報酬は 1 episode で 1 回のみ与えた．

4.3. 距離算出による報酬

3.2 節で挙げた①，③，④の動作の実現を目的として，ある 2 点間の距離を算出し，最短距離を更新した場合に正の報酬(+0.2)，最長距離を更新した場合に負の報酬(-0.2)を与えた(以下，距離更新の報酬と表記する)．アームがワークを掴む前は，先端とワークの距離更新の報酬を与えた．ワークを掴んだ後は，ワークと CP の距離更新の報酬を与えた．CP 通過後は，先端と終箱の距離更新の報酬およびワークと終箱の距離更新の報酬を与えた．なお，先端と終箱間およびワークと終箱間の距離を算出する際には，高さ情報の有無によって場合分けを行った．具体的には，アームが終箱の上部に到達するまでは，高さを除く 2 次元上における 2 点間の距離を，アームが終箱の上部にある場合は，高さを含む 3 次元上における 2 点間の距離を算出した．加えて，アームが終箱の上部にあり，ワークと終箱の最短距離を更新した場合に，ワークと終箱の中心の x，z 座標をそれぞれ取得し，2 点間の距離が一定より小さい場合に 2 段階に分けて報酬(+0.5 もしくは +1.0)を与えた．

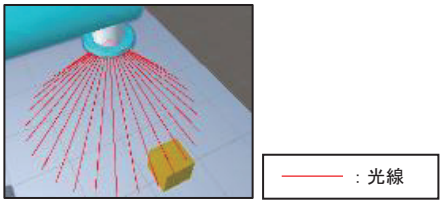

図 2. 光線を射出する様子

5. 評価方法

学習済みモデルの運搬動作における成功率を算出し，評価した．なお，成功率とはワークの生成回数に対して，ワークを終箱に入れた回数とし，成功率算出の際には，衝突発生の有無は考慮していない．

6. 学習結果および考察

step 数の経過による報酬推移のグラフを図 3 に示す．学習が進むのに従って，報酬が増加し，3800 万 step 付近で収束していることがわかる．また，2200 万 step 以降は，報酬の増加があまり認められず，90 前後で報酬が上下している．これは，報酬の大部分が距離更新の報酬であるため，微小な経路の変化が報酬に影響を与えたと考える．

また，学習済みモデルを使用して，成功率を算出した結果，生成回数 10910 回に対して，ワークを終箱に入れた回数は 10901 回であり，成功率は 99.92%である．そのため，5000 万 step の学習の結果，ワークを始箱から終箱に運搬する動作の学習を行うことができたといえる．しかしながら，ワークを終箱に入れる際に衝突が 0.04%の確率で発生した．したがって，先端またはワークと箱との衝突回避を目的とした学習における負の報酬の有用性および効率的な学習法について検討する予定である．

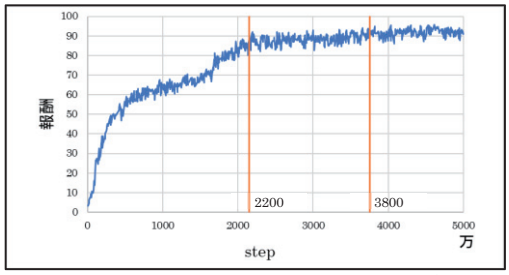

図 3. step 数の経過による報酬推移

参考文献

[1] “World Robotics 2022”, International Federation of Robotics, https://ifr.org/downloads/press2018/2022_WR_extended_version.pdf (Accessed 2023/12/22)
[2] 後藤聡: “産業用ロボットアームの解析と制御に関する研究”, 大阪大学, 博士論文(1995)
[3] “第 3 章 ロボットの普及状況”, 農林水産省, https://www.maff.go.jp/j/budget/yosan_kansi/sikkou/tokutei_keihi/seika_R03/attach/pdf/itaku_R03_ippan-19.pdf (Accessed 2023/12/22)
[4] 原田研介: “産業用ロボットによる組み立て作業の自動化に関する研究動向”, 精密工学会誌, Vol.84, No.4, pp.299-302 (2018)
[5] “Unity”, Unity Technologies, https://unity.com/ja (Accessed 2023/12/22)
[6] “UNITY MACHINE LEARNING AGENTS”, Unity Technologies, https://unity.com/ja/products/machine-learning-agents (Accessed 2023/12/22)
[7] “MATLAB による強化学習”, MathWorks, https://jp.mathworks.com/content/dam/mathworks/ebook/gated/jp-reinforcement-learning-ebook-all-chapters.pdf (Accessed 2023/12/22)
[8] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov: “Proximal Policy Optimization Algorithms”, https://arxiv.org/abs/1707.06347 (Accessed 2023/12/22)
[9] “Ray”, Unity Documentation, https://docs.unity3d.com/ja/2021.3/ScriptReference/Ray.html (Accessed 2023/12/22)

Copyright ©2024 Information Processing Society of Japan.
All Rights Reserved.