モデルベース強化学習による拮抗配置された細径人工筋肉の制御

長友 映里 *† 齊藤 大智 *† Jefferson Pardomuan* 小池 英樹 *

概要. 細径人工筋肉を用いた外骨格グローブは侵襲性が低いため注目されている．細径人工筋肉を用いて手指制御を行うことは高い非線形性やヒステリシスから困難であり，制御手法は未だに確立されていない．本研究では，モデルベース強化学習を用いた細径人工筋肉による軌道追従制御の学習が可能であるかを検証する．一般によく用いられる PID 制御と比較してモデルベース強化学習によって学習された制御則では平均二乗誤差が約 65% 低下することが分かった．

1 はじめに

外骨格グローブはリハビリテーションや技能継承における手指制御の用途で重要な役割を果たしている [1]．その中でも侵襲性が低い細径人工筋肉 [7] を用いた外骨格グローブが注目されている [8]．細径人工筋肉は空気圧で駆動する素材で，軽く柔らかいという性質がある．そのような空気圧駆動の柔らかい素材によって手指制御をするには，高い非線形性やヒステリシスの影響を考慮する必要がある．そのため，細径人工筋肉を用いて手指の動きを再現するような軌道追従制御を行うのは困難であり，制御手法は未だに確立されていない．

細径人工筋肉のような空気圧駆動の柔らかい素材に対する制御手法として，開ループ制御 [9][6] や PID 制御 [3][4] が多く用いられている．手指や細径人工筋肉の動きを定式化することは困難であるため，開ループ制御では精密さに欠ける．PID 制御では手作業によるパラメータ調整が必要である．調整の不要なデータ駆動型手法である深層強化学習を用いた制御手法 [2] も提案されている．Fan ら [2] はモデルフリー強化学習を用いて制御則を学習している．モデルフリー強化学習では膨大な量のデータが必要であるため，一般的にシミュレータを用いてデータを集めながら学習を行う．手指や細径人工筋肉のシミュレータを手に入れることは容易ではないため，細径人工筋肉による手指制御にモデルフリー強化学習を適用することは困難である．そこで，本研究では比較的少量のデータで学習が可能なモデルベース強化学習により制御則を学習することを目指す．

モデルベース強化学習により実機の制御則を学習した研究として，DreamerV2[10] が挙げられる．この研究では細径人工筋肉のような柔らかいロボットへのモデルベース強化学習の適用可能性はまだ調査されていない．そこで本研究では，モデルベース強化学習の細径人工筋肉への適用可能性を調査する．また，本研究で用いるモデルベース強化学習手法の DreamerV2 を軌道追従制御に適用できるように拡張する．

2 手法

本研究では DreamerV2 [5] を用いて軌道追従制御の学習を行った．DreamerV2 では実機で収集したデータにより制御対象のモデルを学習する．そして，制御対象のモデルから生成されたデータを用いることで価値関数と方策を学習する．学習した制御対象のモデルからランダムな目標軌道を生成することはできないため，軌道追従制御に DreamerV2 をそのまま適用することはできない．そこで本研究では，報酬と価値関数，方策を目標軌道で条件付けることで DreamerV2 を軌道追従制御に適用できるように拡張した．拡張した DreamerV2 に含まれる制御対象のモデルと価値関数，方策は以下の式で表される．

$$
\begin{aligned}
h_t &= f_\phi(h_{t-1}, z_{t-1}, a_{t-1}) \\
z_t &\sim q_\phi(z_t|h_t, x_t) \\
\hat{z}_t &\sim p_\phi(\hat{z}_t|h_t) \\
\hat{x}_t &\sim p_\phi(\hat{x}_t|h_t, z_t) \\
\hat{r}_t &\sim p_\phi(\hat{r}_t|h_t, z_t, \tau_{t-1}) \\
\hat{\gamma}_t &\sim p_\phi(\hat{\gamma}_t|h_t, z_t) \\
\hat{a}_t &\sim p_\psi(\hat{a}_t|\hat{z}_t, \tau_t) \\
v_\xi(\hat{z}_t, \tau_t) &\approx E_{p_\phi, p_\psi}\left[\sum_{t \le T} \hat{\gamma}^{T-t} \hat{r}_t\right]
\end{aligned}
$$

z_t が確率的潜在変数，h_t が決定論的潜在変数，a_t が行動，x_t が制御対象から得られる観測，r_t が報酬，γ_t がエピソードが終了したかどうかの二値，τ_t が目標軌道である．予測ステップ長 H だけ生成されたデータを用いて価値関数と方策が学習される．デー

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 東京工業大学
† These authors contributed equally.

タ収集とモデル学習は非同期的に行われるように実装した [11].

本研究では，観測 x_t は各関節の角度，角速度，空気圧から構成される．行動 a_t は細径人工筋肉の空気圧である．目標軌道 τ_t は l ステップ先までの目標関節角度，角速度から構成される．関節の数を n，細径人工筋肉の数を m とすると，$x_t \in \mathbb{R}^{2n+m}$，$a_t \in \mathbb{R}^m$，$\tau_t \in \mathbb{R}^{2ln}$ である．

3 実験

本手法が細径人工筋肉に適用できるのか，またその性能を確認するために，人間の指を模倣した装置で学習を行い，PID 制御との比較を行った．

3.1 設定

中手骨から指骨までの人間の手を模倣した二つのリンクと一つの関節 ($n = 1$) で構成された装置で実験を行った (図 1)．二つのリンクは回転角度を取得するためのロータリーエンコーダによって接続されており，回転角度は ESP32 WROOM 32 に送信される．また，学習に使う目標軌道は 3 ステップ先 ($l = 3$) までの目標関節角度，角速度を使い，報酬は以下のように設定した．

$$r_t = -(\frac{\|\theta_t - \hat{\theta}_t\|}{150})^2 - (\frac{\|\dot{\theta}_t - \hat{\dot{\theta}}_t\|}{7000})^2 - (\frac{p_t^f + p_t^e}{50000})^2$$

θ_t が観測角度，$\hat{\theta}_t$ が目標角度，$\dot{\theta}$ が観測角速度，$\hat{\dot{\theta}}_t$ が目標角速度，p_t^f が屈曲する側の人工筋肉に人工筋肉に送られる空気圧の値，p_t^e が伸展する側の人工筋肉に送られる空気圧の値である．

人工筋肉には s-muscle 社の直径 2.5mm の McKinbben 型の人工筋肉を使用した．人間の筋構造を模倣して屈曲と伸展に関わる 2 本の筋肉を拮抗する形で配置した ($m = 2$)．

学習では $-50°$ から $50°$ の範囲を 0.1 秒ごとに $-10°$ から $10°$ のランダムな値を移動する目標軌道を約 40 分間学習させた．本手法と PID 制御を線形，非線形を含む $10°$ の目標軌道を利用してそれぞれの軌道で 200 ステップずつ試行することで比較した．PID 制御のパラメータは Ziegler-Nichols 法によって概算した後，実験的試行を通じて微調整することによって決定した．

3.2 結果

本手法と PID 制御の平均二乗誤差はそれぞれ表 1 のようになった．そのうち最も結果に差があった軌道は等速に $10°$ ずつ変化する軌道で本手法と PID 制御の平均二乗誤差はそれぞれ 200，1102 となり大きな差が見られた (図 2-a)．反対に，最も結果に差がなかった軌道は 20 ステップごとに $5°$ ずつ移動する軌道で，本手法と PID 制御の平均二乗誤差はそれぞれ 92，46 となり PID 制御の方が優れていた (図 2-b)．学習時には角度変化の小さい滑らかな軌道の学習を行っていないため，PID 制御の方がこの軌道においては良い結果が得られたのではないかと考えられる．

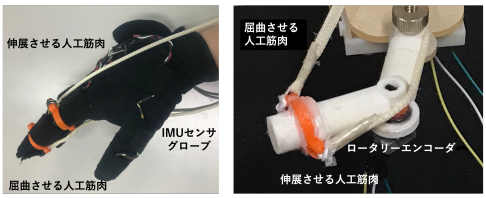

図 1. 左が人間の指に人工筋肉を装着した図，右が実験に使用した人間の指を模倣した装置．

表 1. DreamerV2 と PID 制御による 10 個の線形，非線形を含む同一目標軌道の追従結果 (平均二乗誤差)

	Dreamer-V2	PID 制御
平均二乗誤差	135.7 ± 40.0	386.4 ± 321.1

また，平均二乗誤差の標準偏差は，本手法の方が PID 制御よりも小さくなった (表 1)．このことから本手法は PID 制御よりも多くの軌道に対応できる可能性があることが示された．

4 結論

本研究ではモデルベース強化学習手法である DreamerV2 を軌道追従制御に拡張し，拮抗配置された細径人工筋肉による手指制御の学習を行った．人工筋肉の制御に多く用いられる PID 制御と比較することにより，本手法の有意性と汎化性を示した．今後は細径人工筋肉を人間の指に装着した場合でも同様に軌道追従制御を行うことができるのかを検証する．

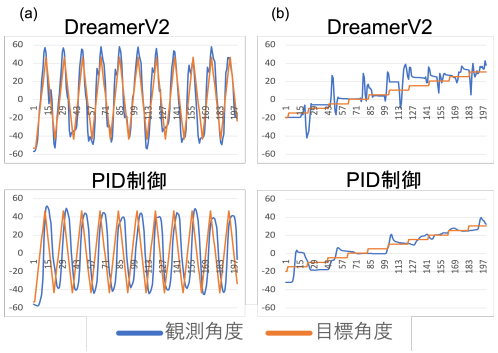

図 2. (a) 最も結果に差があった軌道．(b) 最も結果に差がなかった軌道．上が本手法，下が PID 制御の結果．

謝辞

本研究は JST CREST JPMJCR17A3 の支援を受けたものである．

参考文献

[1] T. du Plessis, K. Djouani, and C. Oosthuizen. A review of active hand exoskeletons for rehabilitation and assistance. *Robotics*, 10(1):40, 2021.

[2] J. Fan, J. Jin, and Q. Wang. Humanoid muscle-skeleton robot arm design and control based on reinforcement learning. In *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 541–546. IEEE, 2020.

[3] A. A. M. Faudzi, N. H. I. M. Lazim, and K. Suzumori. Modeling and force control of thin soft McKibben actuator. *International Journal of Automation Technology*, 10(4):487–493, 2016.

[4] M. H. A. Hafidz, N. M. H. Norsahperi, H. W. Soon, M. N. Jamaludin, and A. A. M. Faudzi. Control of Thin McKibben Muscles in an Antagonistic Pair Configuration. In *Computational Intelligence in Machine Learning*, pp. 155–163. Springer, 2022.

[5] D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba. Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*, 2020.

[6] S. Koizumi, T.-H. Chang, H. Nabae, G. Endo, K. Suzumori, M. Mita, K. Saitoh, K. Hatakeyama, S. Chida, and Y. Shimada. Soft robotic gloves with thin mckibben muscles for hand assist and rehabilitation. In *2020 IEEE/SICE International Symposium on System Integration (SII)*, pp. 93–98. IEEE, 2020.

[7] S. Kurumaya, H. Nabae, G. Endo, and K. Suzumori. Design of thin McKibben muscle and multifilament structure. *Sensors and Actuators A: Physical*, 261:66–74, 2017.

[8] N. Takahashi, S. Furuya, and H. Koike. Soft exoskeleton glove with human anatomical architecture: production of dexterous finger movements and skillful piano performance. *IEEE Transactions on Haptics*, 13(4):679–690, 2020.

[9] B. Wang, N. Takahashi, and H. Koike. Sensor Glove Implemented with Artificial Muscle Set for Hand Rehabilitation. In *Proceedings of the Augmented Humans International Conference*, pp. 1–4, 2020.

[10] P. Wu, A. Escontrela, D. Hafner, P. Abbeel, and K. Goldberg. DayDreamer: World Models for Physical Robot Learning. In *6th Annual Conference on Robot Learning*.

[11] Y. Zhang, I. Clavera, B. Tsai, and P. Abbeel. Asynchronous Methods for Model-Based Reinforcement Learning. In *Conference on Robot Learning*, pp. 1338–1347. PMLR, 2020.

未来ビジョン

外骨格グローブはリハビリテーションや技能継承等の幅広い用途への応用が期待されている．その中でも空気圧駆動の細径人工筋肉を用いたグローブは侵襲性が低く人間の手に装着するのに相性が良い．また，技能獲得において重いグローブを用いた場合，練習時と実践時で重さに関する感覚に乖離が出てしまい練習効率が下がる可能性がある．そのような観点でも，軽い細径人工筋肉を用いたグローブは期待されている．しかしながら細径人工筋肉の強い非線形性やヒステリシスによる制御の難しさから制御手法が確立されておらず，その用途は限られていた．本研究では，人間の手を模倣した装置を使ってモデルベース強化学習による制御の学習を行い，その有効性を検証した．

今後は，細径人工筋肉を人間の指に装着して本手法による制御を行う予定である．本研究では関節の数を 1 として制御を行ったが，実際の指の関節の数は 3 以上あるため関節の数に関して拡張する．さらに，手指の筋肉の状態によって指を動かすために必要な空気圧が変わってしまう Human-Robot Interaction の問題にも対処する予定である．筋肉の状態を学習に加えることによってさらに正確な制御が可能になるのではないかと考えている．筋肉の状態をフィードバックとして含めることは，制御値の誤差を用いて操作値を更新するような PID 制御ではできない．本手法のような深層学習を用いた手法では制御値以外の複数モダリティの統合が可能であり，そういった点でも PID 制御より優れているのではないかと考えている．

最終的には複数の指を制御することによって，外骨格グローブのリハビリテーションや技能継承への効果・効率を上げること目標としている．