深層強化学習を用いた二脚モデルにおける歩容遷移の実現

○古関駿介（東北大） 沓澤京（東北大）大脇大（東北大）林部充宏（東北大）

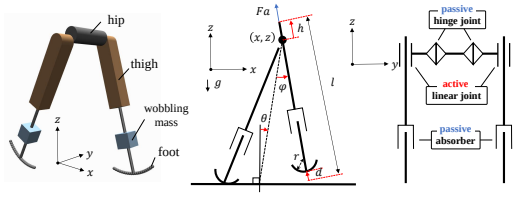
1. 緒言

二脚ロボットにおいて、幅広い速度でエネルギー効率の高い移動を実現させるための 1 つの手段は、歩容を遷移させることである。移動速度に応じてロボットの歩容を歩行から走行に遷移させることで、歩行の上限速度を超えた移動が実現可能となる [1, 2]。例えば、ヒト型ロボット ASIMO が歩行する際の上限速度は約 1.3 m/s であるが、走行に歩容を遷移させることによって最大約 2.5 m/s までの速度で移動可能となる [2]。また、低速度領域では歩行、高速度領域では走行が、エネルギー効率的に最適な運動となることが知られており [3, 4]、歩容を遷移させることで、高いエネルギー効率の運動を維持することも可能となる。

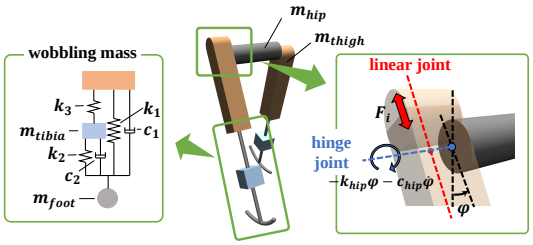
ヒトは、身体の力学的特性を巧みに活用することで、歩行と走行間の遷移を実現している [5]。歩行運動において、支持脚の運動は倒立振子、遊脚の運動は腰を視点とする振り子として力学的にモデル化され [6]、力学的エネルギーが一定となるよう運動エネルギーと位置エネルギーを互いに変換しながら移動している。この原理は、重力を活用し緩やかな斜面を歩き下る受動歩行機械 [7] の歩容生成においても確認できる。一方、走行運動では、バネが実装された倒立振子モデル（Spring Loaded Inverted Pendulum: SLIP）[8, 9] としてモデル化され、脚に存在する筋や健の協働的な効果から弾性エネルギーを活用した運動が生成されている。このように、ヒトの歩行運動と走行運動は、身体特性に基づく力学的に異なった運動モードとして生成されていると考えられる。しかしながら、歩容遷移を試みた二脚ロボット [10, 11] において、身体のダイナミクスを有効に活用した研究はほとんどない。

二脚ロボット制御において、身体ダイナミクスを活用する歩容遷移が困難な理由は以下の 2 点である：(1) 二脚ロボットでは、本質的に劣駆動系となるため、全てのアクチュエータを駆動するロボットと比べ制御は困難となる。(2) 上述のように歩行運動と走行運動は力学的に異なる運動モードと考えられるため、単一のコントローラにより複数の運動モードを生成することは難しい [12]。この問題を解決するため、本研究では、深層強化学習に注目する。四脚ロボットにおいて，深層強化学習を用いることで、異なる歩容を生成する制御器を学習できることが示唆されている [13]。

本研究では、深層強化学習を用いて、身体ダイナミクスを活用する二脚モデルの歩容遷移を再現することを目的とする。この目的のため、股関節が受動的なモデルを構築し、深層強化学習の報酬関数と学習方針の設計、生成された歩容を評価した。その結果、単一のコントローラによって、1 つのパラメータのみの変化で歩行と走行を再現することに成功した。



(a) 左：モデル　中：side view　右：front view



(b) Wobbling 機構、股関節の構造

図 1　使用した二脚モデル

2. 提案手法

2.1 二脚モデル

受動歩行モデル [14] を参考にして作成した、本研究で用いる二脚モデルを図 1 に、身体パラメータを表 1 に示す。運動は図 1(a) の xz 面（矢状面）に拘束されている。本モデルの特徴は、股の回転関節が完全に受動的であるため、脚を前後に振るためにアクチュエータではなく、重力による位置エネルギーとばねによる弾性エネルギーを使用する点である。これにより、積極的に身体のダイナミクスを活用する移動運動の生成が見込める。各脚には最大出力が F のリニアアクチュエータが取り付けられており、脚方向 (図 1 (b) 右の赤点線方向) に沿って大腿部を上下させる。これらのアクチュエータは、支持脚前期に地面を押して推進力を発生させ、支持脚後期に脚を持ち上げて足底と地面の衝突を防ぐ働きを持つ。股関節 (図 1(b) の青点線方向) は、巻きバネ（k_{hip}）とダンパー（c_{hip}）を持つ受動的な回転関節とした。k_{hip} は股関節に脚を前後に振る受動的な回転力を発生させ、さらに脚が大きく開きすぎるのを防ぐ。また、下腿部は 3 つの線形ばね（k_1、k_2、k_3）、2 つのダンパー（c_1、 c_2）、質量（m_{tibia}）からなる wobbling 機構を採用した（図 1(b)）[15]。この機構により、着地時の足裏と地面との衝撃を低減する効果が得られる。足底は、20 個の小さな球から構成される曲率半径 r の円弧形状とした。モデルの状態は、腰部の位置と向き (x, z, θ)、左右脚（左：j=l、右：j=r）の角度 (φ_j)、大腿部の変位 (h_j)、脚の収縮量 (d_j) に

表 1　二脚モデルの身体パラメータ

	Unit	Value
l	[m]	0.8
r	[m]	0.27
m_{hip}	[kg]	20
m_{thigh}	[kg]	6
m_{tibia}	[kg]	3
m_{foot}	[kg]	1
k_{hip}	[Nm/rad]	25
c_{hip}	[Nms/rad]	2
k_1	[N/m]	6000
k_2	[N/m]	6000
k_3	[N/m]	10000
c_1	[Ns/m]	300
c_2	[Ns/m]	650
F	[N]	600

より与えられる。

2.2　深層強化学習

本研究では、深層強化学習に連続制御タスクのための最先端のアルゴリズムである Soft Actor-critic (SAC) を採用した。このアルゴリズムでは、ボーナス報酬として方策のエントロピー $H(\pi)$ に比例した値が加えられることが特徴である。以下の目的関数 $J(\pi)$ を最大化する確率的な方策 π を求める。

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}[\gamma^t (r(s_t, a_t) + \alpha H(\pi(\cdot|s_t)))], \quad (1)$$

ここで，γ は割引率、r は後述する報酬関数、α は温度パラメータと呼ばれ、エントロピー項の重視度を決定する。s_t と a_t はそれぞれ状態と行動を表す。

2.3　学習方法

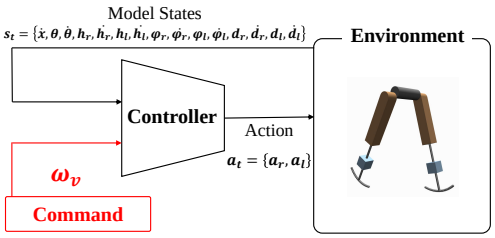
本研究では、入力される速度を規定するコマンド ω_v に従って、出力（歩行と走行）を変えるコントローラーの獲得を目指す。学習は「学習期間 I」と「学習期間 II」の二段階に分けた。学習期間 I では低速度運動を学習し、学習期間 II では幅広い速度の移動運動の学習を行う。学習後は、図 2 のように一つのパラメータ ω_v のみでモデルの歩容を制御することを目指す。

2.3.1　報酬関数

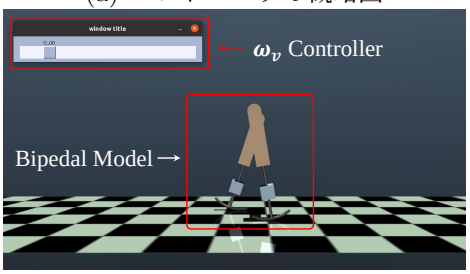
どちらの学習期間でも、以下の報酬関数 r を使用する。ハイパーパラメータの値は表 2 の通りである。

$$\begin{aligned} r(s_t, a_t) = &- \omega_E |\Delta E_t| + \omega_v \dot{x} \\ &+ f_{forward} + f_{alive} + f_{support}, \end{aligned} \quad (2)$$

$$f_{forward} = \begin{cases} 0 & (\dot{x} \geq 0) \\ -C_1 & (\dot{x} < 0) \end{cases}, \quad (3)$$



(a) コントローラの概略図



(b) シミュレーション環境上の制御の様子

図 2　使用したコントローラ

表 2　報酬関数と SAC のパラメータ

Reward function						SAC	
ω_E	ω_l	ω_s	C_1	C_2	C_3	α	γ
0.2	0.2	0.15	1.0	1.0	0.5	0.2	0.99

$$f_{alive} = C_2, \quad (4)$$

ここで、ω は重み係数、ΔE_t は t ステップ時のモデルの総エネルギー (運動、位置、弾性エネルギーの和) 変化、$\dot{x}$ は hip セグメントの x 方向速度である。第一項により、エネルギー変化が大きい運動であるほど負の報酬が与えられる。第二項により、モデルの速度が大きいほど正の報酬が与えられる。第三項 $f_{forward}$ はモデルを常に前進させるための項である。モデルが後退、つまり $\dot{x} < 0$ のとき負の報酬が与えられる。第四項 f_{alive} はモデルの転倒を押さえる項である。本研究では、$|\theta| > 1.4[rad]$、つまりモデルが 80 度以上傾くと、モデルは転倒しているとみなし、上限ステップに達していなくても、エピソードの終わりとした。モデルが倒れなければこの項により正の報酬が常に与えられ、累積報酬を大きくなる。第五項 $f_{support}$ は学習を効率的に行うための補助的な項である。

$$f_{support} = \omega_l f_{leg} + \omega_s f_{sym}, \quad (5)$$

$$f_{leg} = \min(|\dot{\varphi}_r - \dot{\varphi}_l|, C_3), \quad (6)$$

$$f_{sym} = -|\pi(s_t) - \Psi_a(\pi(\Psi_o(s_t)))|^2, \quad (7)$$

ここで、添え字の r と l はそれぞれ右脚と左脚を示す。f_{leg} は脚を前後に振る動きを促進する項であり、φ は脚の開いている角度である。前述したように、モデルは脚を直接前後に振ることができないため、この項を設定した。f_{sym} は、方策に左右対称性を与える項である [16]。Ψ_a、Ψ_o はそれぞれ a_t と s_t を鏡対称に写像する関数である。現在の状態と鏡の中の状態が起こす状態を比較し、出力が異なるほど負の報酬を与える。

2.3.2 学習期間

本研究の学習の特徴は、速度の重みである ω_v の値が 1 エピソードごとにランダムで変化する点である。速度を規定する ω_v に対する適切な出力を学習することで、歩行と走行の発現を目指す。そのため、学習時に式 (1) に入力する s_t には、二脚モデルの状態に加えて、速度項の重み係数 ω_v を与える。具体的には、$s_t = \{\dot{x}, \theta, \dot{\theta}, h_r, \dot{h}_r, h_l, \dot{h}_l, \varphi_r, \dot{\varphi}_r, \varphi_l, \dot{\varphi}_l, \dot{d}_r, \dot{d}_l, \omega_v\}$ である。

学習期間 I と学習期間 II の違いは、ω_v の範囲である。学習期間 I では $\omega_v \in [-0.2, 0.2]$、学習期間 II では $\omega_v \in [-0.5, 3.5]$ の範囲で値がランダムに設定する。これにより，学習期間 I では、式 (2) より、速度を重視しないエネルギー変化の小さい前に進む運動を獲得する傾向が高く、学習期間 II では、幅広い速度の重視度をエピソードごとに変えて運動を学習する傾向が高い。

3. 結果

学習は、10,000,000 ステップ行った。初めの 4,500,000 ステップは学習期間 I、残りは学習期間 II とした。また，学習期間 I の初期 500,000 ステップでは、二脚モデルが適切な歩容を学習できるよう、重み係数 ω_E と，ω_s をそれぞれ $\omega_E = 0.06$、$\omega_s = 0$ とすることで学習時の制約を緩くした。

3.1 歩容遷移実験

図 3 の上図は、コントローラに与える ω_v の値を、0 から 3.0 へ 0.4 秒間で連続的に増加させたときのエネルギーと速度の時間変化である。図中の斜線部分が ω_v を変化させている期間である。下図は、ω_v の値を、3.0 から 0 に減少させたときの様子である。両脚支持期を持つ歩容を歩行、両脚遊脚期を持つ歩容を走行とすると [17]、図 3 より $\omega_v = 0$ のとき両脚支持期（紫の期間）が見られることから歩行が、さらに $\omega_v = 3.0$ のとき両脚遊脚期（白の期間）が確認できることから走行が発現していることがわかる。

また、歩行中は運動エネルギーと位置エネルギーが逆相であることが、走行中は片脚支持期に運動エネルギーと位置エネルギーが同相であり弾性エネルギーがこれらと逆位相になっていることがわかる。これらは、歩行と走行で確認される特徴であることから [5]、エネルギー変化の観点から見ても歩行と走行が生成されているといえる。

3.2 学習期間が与える影響

学習期間 I がある場合とない場合の学習曲線を図 4 に示す。累計報酬から判断すると、学習期間 I があるとき学習が効率よく行われていることがわかる。

4. 考察

生成された歩容、エネルギー変化より、深層強化学習を用いて二脚モデルの歩容遷移を実現できたと結論できる。報酬関数には式 (2) を用いた。この中でも第一項 $\omega_E|\Delta E_t|$ が歩容生成に大きな役割を果たしていると考える。この項がないとモデルは前進するが、脚を引きずる上に不自然に跳ね、定常的な歩行と走行は生成される様子は観察されなかった。ヒトは身体を生かし、歩行では倒立振り子のように運動、位置エネルギーを

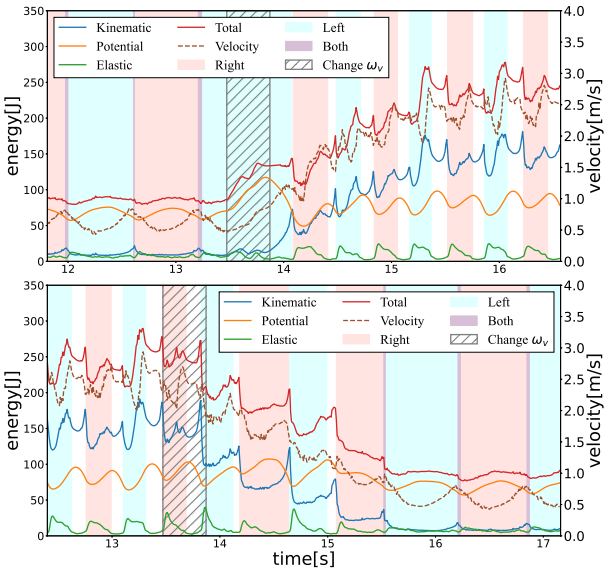


図 3 ω_v を変えたときの二脚モデルのエネルギーと速度の変化。上図が $\omega_v : 0 \to 3.0$（歩行 → 走行）。下図が $\omega_v : 3.0 \to 0$（走行 → 歩行）

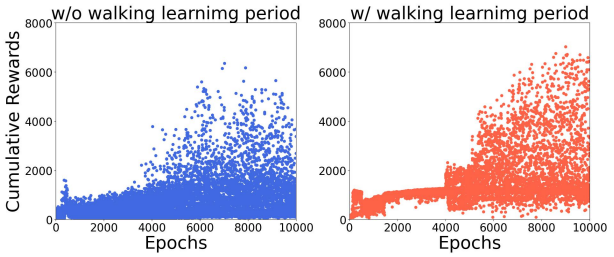


図 4 歩行学習期間の有無での累積報酬比較
左が歩行学習期間なし、右が歩行学習期間あり。

交換することで、走行では運動、位置エネルギーの一部を弾性エネルギーとして保存することでエネルギー変化を抑え高効率な移動を行っている [5]。ヒトのように身体を活用し、移動運動時のエネルギー変化を抑えたことで、歩行と走行が再現されたと考える。

また、初めから幅広い速度の運動を同時に学習させるのではなく、学習初期に低速度の運動のみを学習させると、学習の成功率が高いことが示唆された（図 4）。学習期間 I がないとき、ある範囲の ω_v で走行を学習することがあったが、図 4 左で累計報酬が 0 付近に密集していることが示すように、入力されたほとんどの ω_v で転倒してしまった。一方、図 4 右では、累計報酬が 0 付近にほぼないことから、学習期間 I を経ることで、入力される ω_v に対応する歩容が生成されたことがわかる。この結果は、幅広い速度の運動を同時に学習させる場合、学習初期に低速度運動の学習を経ることで効率的な運動学習が可能となることを示唆する。

生成した歩容のエネルギー効率を以下に定義する CoT (Cost of Transport) により導出する。

$$CoT = \frac{1}{mg\Delta d} \sum_{i \in r,l} \int_{t_0}^{t_{end}} \max(Fa_i \dot{h}_i, 0) dt, \qquad (8)$$

ここで、m はモデルの質量、Δd は移動距離、g は重力

加速度、Fa_i はアクチュエータの出力、$\dot{h}_i$ は大腿部の速度である。CoT が小さいほど、エネルギー効率がよい。生成した歩容の CoT は、歩行で 0.342 ($\omega_v = 0$、$Fr = 0.22$)、走行で 0.347 ($\omega_v = 3.0$、$Fr = 0.93$) であった。ここで、$Fr = \frac{v}{\sqrt{gl}}$ はフルード数と呼ばれ、速度を脚長により正規化した値である。参考として、歩容遷移を行うことが可能である ASIMO の歩行時の CoT は $= 3.2$（$Fr = 0.184$）[18] である。

次に、速度を先行研究と比較する。生成した歩容は、$Fr = 0.16 \sim 0.93$ で移動が可能であった。小林らの先行研究は約 $Fr = 0 \sim 0.67(l = 0.595)$、ASIMO は約 $Fr = 0 \sim 1.02(l = 0.61)$ で移動をしていた。我々の歩容は、先行研究と同等の最大速度での移動が可能であり、幅広い移動速度を持つことがわかる。一方で、本研究で使用したモデルは、先行研究と異なり、股関節 [19] が受動的である。股関節の駆動は、ヒトが速度によって大殿筋、大腿直筋などの活動度を変えるように [20]、移動速度に大きな影響を与える [19]。よって、本研究の結果は、股関節が能動的でなくても、身体のダイナミクスを有効に活用することで、二脚モデルに幅広い速度域の運動を生成できることを示唆する。

5. 結言

本研究では、深層強化学習を用いて身体のダイナミクスを活用する二脚ロボットの歩容遷移を再現した。得られたコントローラは速度を規定するパラメータ ω_v のみを変化させるだけで、歩容遷移を再現することに成功した。本研究の結果は、歩容遷移により幅広い速度領域でエネルギー効率の高い移動を可能とする二脚ロボットを開発するための枠組みになると期待する。

参 考 文 献

[1] Koushil Sreenath, Hae-Won Park, and J. W. Grizzle. Design and experimental implementation of a compliant hybrid zero dynamics controller with active force control for running on mabel. In *2012 IEEE International Conference on Robotics and Automation*, pp. 51–56, 2012.

[2] HONDA 社 ASIMO のウェブサイト. `https://www.honda.co.jp/ASIMO/`.

[3] Manoj Srinivasan and Andy Ruina. Computer optimization of a minimal biped model discovers walking and running. *Nature*, Vol. 439, No. 7072, pp. 72–75, 2006.

[4] Frederick J Diedrich and William H Warren Jr. Why change gaits? dynamics of the walk-run transition. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 21, No. 1, p. 183, 1995.

[5] GA Cavagna and MJTJop Kaneko. Mechanical work and efficiency in level walking and running. *The Journal of physiology*, Vol. 268, No. 2, pp. 467–481, 1977.

[6] Arthur D. Kuo. The six determinants of gait and the inverted pendulum analogy: A dynamic walking perspective. *Human Movement Science*, Vol. 26, No. 4, pp. 617–656, 2007.

[7] Tad McGeer. Passive dynamic walking. *Int. J. Robotics Res.*, Vol. 9, No. 2, pp. 62–82, 1990.

[8] Reinhard Blickhan. The spring-mass model for running and hopping. *Journal of biomechanics*, Vol. 22, No. 11-12, pp. 1217–1227, 1989.

[9] Michael H Dickinson, Claire T Farley, Robert J Full, MAR Koehl, Rodger Kram, and Steven Lehman. How animals move: an integrative view. *science*, Vol. 288, No. 5463, pp. 100–106, 2000.

[10] Kenichiro Nagasaka, Yoshihiro Kuroki, Shinya Suzuki, Yoshihiro Itoh, and Jinichi Yamaguchi. Integrated motion control for walking, jumping and running on a small bipedal entertainment robot. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, Vol. 4, pp. 3189–3194. IEEE, 2004.

[11] Taisuke Kobayashi, Yasuhisa Hasegawa, Kosuke Sekiyama, Tadayoshi Aoyama, and Toshio Fukuda. Unified bipedal gait for walking and running by dynamics-based virtual holonomic constraint in pdac. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1769–1775, 2016.

[12] Nils Smit-Anseeuw, Rodney Gleason, Ram Vasudevan, and C David Remy. The energetic benefit of robotic gait selection—a case study on the robot ramone. *IEEE Robotics and Automation Letters*, Vol. 2, No. 2, pp. 1124–1131, 2017.

[13] Zipeng Fu, Ashish Kumar, Jitendra Malik, and Deepak Pathak. Minimizing energy consumption leads to the emergence of gaits in legged robots. 10 2021.

[14] Dai Owaki, Koichi Osuka, and Akio Ishiguro. On the embodiment that enables passive dynamic bipedal running. In *2008 IEEE International Conference on Robotics and Automation*, pp. 341–346, 2008.

[15] Ali Asadi Nikooyan and Amir Abbas Zadpoor. Mass-spring-damper modelling of the human body to study running and hopping–an overview. *Proceedings of the institution of mechanical engineers, Part H: Journal of engineering in medicine*, Vol. 225, No. 12, pp. 1121–1135, 2011.

[16] Wenhao Yu, Greg Turk, and C Karen Liu. Learning symmetric and low-energy locomotion. *ACM Transactions on Graphics (TOG)*, Vol. 37, No. 4, pp. 1–12, 2018.

[17] R. McNeill Alexander. Walking and running. *The Mathematical Gazette*, Vol. 80, No. 488, p. 262–266, 1996.

[18] Steven H Collins and Andy Ruina. A bipedal walking robot with efficient and human-like gait. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pp. 1983–1988, 2005.

[19] Joshua P Bailey, Tiffany Mata, and John Mercer. Is the relationship between stride length, frequency, and velocity influenced by running on a treadmill or overground? *International Journal of Exercise Science*, Vol. 10, pp. 1067–1075, 2017.

[20] Poppele RE Lacquaniti F Cappellini G, Ivanenko YP. Motor patterns in human walking and running. *Journal of neurophysiology*, pp. 229–235, 2005.