

深層強化学習を用いた準受動歩行および走行の実現

Realization of Quasi-Passive walking and Running Using Deep Reinforcement Learning

○ 古関駿介, 沓澤京, 大脇大, 林部充宏,

○ Shunsuke Koseki, Kyo Kutsuzawa, Dai Owaki, Mitsuhiro Hayashibe

東北大学

Tohoku University

キーワード: 受動歩行(Passive Walking), 深層強化学習(Deep Reinforcement Learning), 歩容(Gait), 二脚ロボット(Biped Robot), エネルギー効率(Energy Efficiency)

連絡先: 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-01 東北大学 工学部 機械知能・航空工学科 林部研究室
古関駿介, Tel.: (022)795-6970, Fax.: (022)795-6970, E-mail: shunsuke.koseki.q4@dc.tohoku.ac.jp

1 緒言

ヒトは, 歩行, 走行と大きく分けて2つの歩容で移動を行っている. 歩行と走行を使い分ける利点の1つに, 幅広い速度で移動を行えることが挙げられる. 一般的な歩行では約3.0[m/s]以上の速度が出せないと言われていたが¹⁾, 走行に歩容を切り替えることにことでより速い速度域での移動が可能となる. また, 速度に応じ消費エネルギーの少ない歩容を選択し, エネルギー効率のよい移動を可能としているとも考えられている (ここでのエネルギーとは, 単位質量を単位距離移動させるのに必要なエネルギー[cal/kg·m]を指す). 前述の通り, 歩行では約3.0[m/s]までの速度を出すことができると言われているが, 実際にヒトが歩行から走行へと歩容を変える速度は約2.1[m/s]であることが知られている. ヒトの運動データから, 移動速度がおおよそ2.2~2.3[m/s]より遅いときには歩行, 速いときには走行を選択すればエネルギー効率

がよいことが示されており, ヒトの歩容が変わる速度とよく一致している. このようなヒトと同様の歩行および走行運動をロボットで再現することができれば, エネルギー効率がよく幅広い速度で移動することのできる二足ロボットの実現が期待できる.

ロボットの分野では, ヒトのような自然な歩容を再現しエネルギー効率のよいロボットを実現するために受動歩行が注目されている²⁾. 受動歩行とは, アクチュエータなどの駆動機構を持たない完全に受動的な歩行機械が, 適切な初期値さえ与えられれば, 緩やかな斜面を安定的に歩き下る現象のことを指す^{2),3)}. 重力による位置エネルギーを運動エネルギーに変換することで移動を行い, 実現される歩容は自然かつ滑らかであり, ヒトの歩行に似ている. ヒトは歩行時, 振子のように前方に倒れ込む動作を繰り返すことで巧みに重力場を使用し, 高いエネルギー効率で移動していることが知られているが, 受動歩

行はその最も単純かつ本質的な形である。一方、平地では重力場からエネルギーを継続的に得ることができないため、完全に受動的な移動はできない。そこで、簡単なアクチュエータを代替のエネルギー供給手段として付け足すことにより、斜面以外での歩行を目指した準受動歩行に関する研究も行われている。例えば、足首にモータをつけることにより地面を蹴り出す力を与える方法、地面から支持脚先端にインパルス入力を与える方法、支持脚方向に出力する簡単な直動アクチュエータを付ける方法、といった手段が提案されている^{4),5),6)}。

このように、準受動歩行を実現させた研究は数多くある。しかし、ヒトは歩行だけでなく、走行で移動することがあるにもかかわらず、実現されてきた準受動的な歩容は歩行のみにとどまっている。ヒトのような移動をロボットで再現するために走行の実現は欠かせないが、準受動歩行と準受動走行を同時に実現させた研究は筆者の知る限り存在しない。

そこで、本研究では同一のモデルで準受動歩行および走行を実現させるための手法を提案する。具体的には、深層強化学習を用いて準受動歩行モデルの持つアクチュエータの出力を学習させることで歩行と走行それぞれの発現を目指す。深層強化学習で用いる報酬関数をどのように設計し、学習を行えば歩行と走行が実現できるかを検証し、それぞれの歩容で身体の使い方によどのような違いがあるかを比較した。

2 提案手法

2.1 歩行と走行の定義

Fig.1 に歩行と走行の移り変わりの様子を示す。歩行では図中の紫色の線で示すように、両脚が地面に接地している両脚支持期と呼ばれる期間がある。一方、走行では図中の緑色の線で見られるように、どちらの脚も空中に浮いている両脚遊脚期と呼ばれる期間がある^{7),8)}。本研究では両脚支持期が周期的に存在する歩容を歩行、両脚遊脚期が周期的に

存在する歩容を走行と定義する。

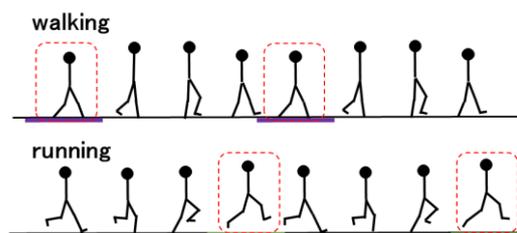


Fig. 1 歩行と走行の移り変わり

2.2 準受動歩行モデル

ヒトは走行中、位置エネルギーと運動エネルギーの一部を手足や胴の弾性要素に保存することで、エネルギーの損失を抑えている⁹⁾。このことから、走行には身体に弾性要素を持たせることが重要とわかる。実際に、股関節に巻きばね(Fig.2 オレンジ色のらせん部分)と、支持脚に直列につながれた直動のばね(Fig.2. 赤色の部分)を持つ、ばね付き受動歩行モデルを用いることで受動的な走行を実現できることが報告されている¹⁰⁾。このばね付き受動歩行モデルを参考に、簡単なアクチュエータを搭載したばね付き準受動歩行モデルを用いて歩行、走行の再現を試みた。

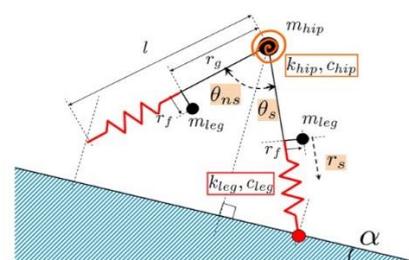


Fig. 2 参考にしたばね付き受動歩行モデル¹⁰⁾

Fig.3 に本研究で用いたばね付き準受動歩行モデルを示す。また、部位のパラメータは table1 の通りである。モデルには脚方向(Fig.3(c)の赤点線)に沿って脚を上下させる直動アクチュエータが搭載されている。このアクチュエータを搭載することで、脚を地面に押しつけることにより蹴り出す力、脚を地

面に衝突させず前に運ぶための持ち上げる力を生むことを期待する。受動的な脚のふりを実現させるために付けた腰関節の巻きばねは、脚が地面に対して垂直なとき (Fig.3(c), $\varphi = 0$) が平衡点であり, Fig.3(c)の青点線周りにトルク $k_{hip}\varphi$ を与え, 脚を前後に振らすことができる。

また, 直動ばね k_2, k_3 に m_{tibia} を挟んだ揺動慣性と呼ばれる部品 (Fig.2(b)) を実装した。揺動慣性をつけることで, モデルが床に衝突したとき, 柔らかい着地をすることができる^{11),12)}。Fig4 はモデルを高さ 5m から速度 5[m/s]で地面に向かって落としたときの床反力である。揺動慣性を付けたとき (赤) は付けていないとき (青) と比べ, 衝突したときに受ける最大床反力が小さくなっていることがわかる。揺動慣性を付けることによって, モデルが歩行および走行中に弾んでしまう問題の解決が期待できる。

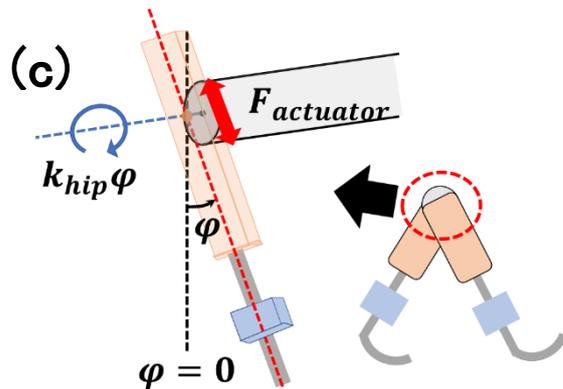


Fig. 3 本研究で使用した巻きばね付き準受動歩行モデル(a)シミュレーション環境上, (b)機構図, (c)ばねとアクチュエータの働き方

table 1 モデルのパラメータ

m_{hip}	10	[kg]	k_1	6	[kN/m]
m_{thigh}	6	[kg]	k_2	6	[kN/m]
m_{tibia}	3	[kg]	k_3	10	[kN/m]
m_{foot}	1	[kg]	c_1	300	[Ns/m]
k_{hip}	87	[Nm/rad]	c_2	650	[Ns/m]
c_{hip}	0.2	[Nms/rad]	$F_{actuator}$	$ F \leq 400$	[N]

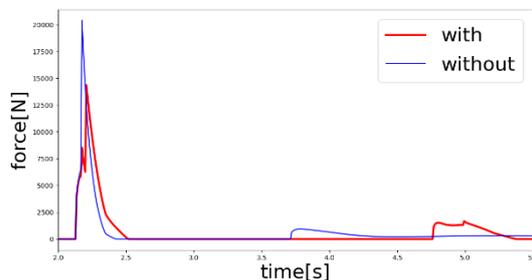
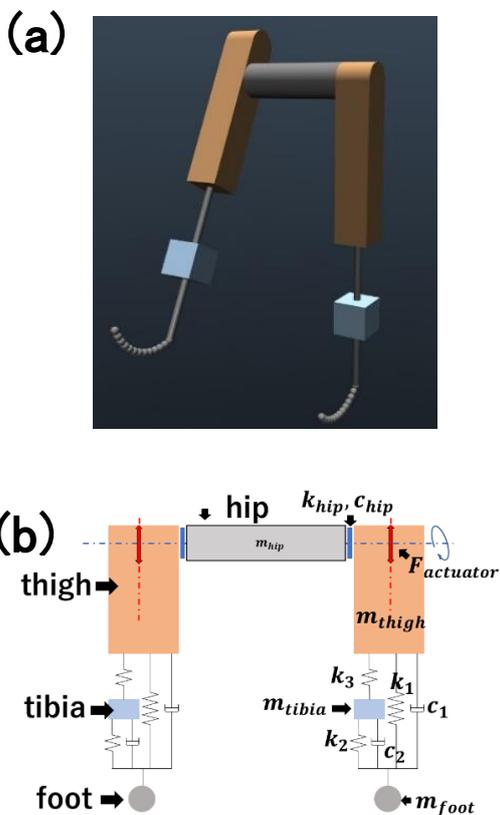


Fig. 4 m_{tibia} の効果

2.3 深層強化学習

強化学習はエージェント (学習者) と, エージェントが置かれた環境との相互作用によって進行する。エージェントは(1)環境から状態

s_t を観測し、(2) s_t に基づいて行動 a_t を起こし、(3)その結果として状態 s_{t+1} へ遷移する共に報酬 r_t を受け取る、と(1)~(3)のサイクルを繰り返しながらタスクを実行する。ここで、(2)において、現在の状態から次の行動を導き出すエージェントが持つ規則を方策 π と呼ぶ。強化学習とは(1)~(3)のサイクルを繰り返すことで得られる経験から、タスク終了までに累計報酬を大きくする方策 π を学習する枠組みである¹³⁾。

強化学習では、最適な方策を決定する指標に状態価値関数 $V(s_t)$ や行動価値関数 $Q(s_t, a_t)$ などの関数が用いられる。しかし、取り得る状態 s 、行動 a が多いとき(例えば連続値の場合)、価値関数を求めるには計算量が多すぎるため現実的ではない。そこで、ニューラルネットワークを用いて価値関数を近似する手段がある。深層強化学習とは、このように強化学習中の関数にニューラルネットワークを用いる手法のことである。

本研究では、深層強化学習のアルゴリズムにSAC(Soft Actor-Critic)を使用した。SACの特徴は、報酬に加えてそのときの方策 π のエントロピー $\mathcal{H}(\pi)$ に比例したボーナス報酬を得ることのできる点である。エントロピー項を加えることにより、広範囲の探索が可能となるため局所解に陥りにくくなる、ロバスト性のある方策を得やすくなる、といったメリットがある¹⁴⁾。式(2.1)で示す目的関数 $J(\pi)$ を最大にする確率的方策 π を学習する。

$$J(\pi) = \sum_{t=0}^T E_{(s_t, a_t) \sim \rho_\pi} \left[\gamma^t \left(r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot | s_t)) \right) \right] \quad (2.1)$$

ここで、 α は温度パラメータと呼ばれ、報酬に対し方策のエントロピーをどの程度重視するかを決める。

2.4 報酬設計

歩行と走行を学習させるために設計した深層強化学習の報酬関数 r について説明する。報酬関数の設計には(Ⅰ)~(Ⅲ)の3つの方針を立てた。

- (Ⅰ)エネルギーの消費を抑えた安定した運動を獲得させる
- (Ⅱ)歩行か走行か、歩容を決定する項を導入する
- (Ⅲ)効率よく学習をさせる

以上の方針のもと設計した報酬関数 r を式(2.2)に示す。(Ⅱ)をもとに設定した項が歩容を決定しているかを知るために、歩行用と走行用の報酬関数は、 ω_2 の値のみが異なり、他は全て同じである。

$$r = -\omega_1 |E_t - E_{t-1}| + \omega_2 v_{model} + assist \quad (2.2)$$

ここで ω は重み係数と呼ばれ、項の数値のバランスを調整することや、報酬関数の中での重視度を定めることができる。

第1項は(Ⅰ)エネルギー消費を抑え、安定した運動を獲得させるために設定した項である。 E_t は t ステップ時におけるモデルの総エネルギー(位置エネルギー+運動エネルギー+弾性エネルギー)である。この項によってモデルの総エネルギーの変化が大きいほど負の報酬が与えられる。モデルが運動中に失ったエネルギーはアクチュエータの仕事により補充をしなければならないので、総エネルギーの変化が大きいほどエネルギー消費が大きくなる。よって、この項によりエネルギー消費を抑えた運動の学習が期待できる。さらにエネルギー変化が小さい行動を学習させることにより、安定した運動の獲得も望める。

第2項は(Ⅱ)歩行か走行か、歩容を決定するための項であり、速度の重視度を定める。 v_{model} はモデルの速度であり、この項によって速度が大きいほどもらえる報酬が大きくなる。重み係数 ω_2 が大きいときは、報酬関数の中で速度の重視度が高まるので、エージェントはより速度を出すことに重点をおいた運動を学習すると考えられる。 ω_2 を小さくすることで歩行、大きくすることで走行の発現が期待できる。

第3項は、(Ⅲ)学習を効率的に行うために

設定した項である．詳細を式(2.3)に述べる．

$$\text{assist} = \omega_3 |\dot{\theta}| - \omega_4 \text{symmetry} + \text{alive bonus} \quad (2.3)$$

第1項の $\dot{\theta}$ は右脚と左脚の角度差(Fig.2の $\theta_s + \theta_{ns}$ にあたる)の時間微分である．股関節の巻きばねによって脚を振る動きの学習が困難であったため設定した．第2項は，方策に左右対称性を与える項である¹⁵⁾．ヒトは正常歩行および走行時，周期的に見て左右の動きが対称になるので，この項を設定した．第3項は，モデルが転ばないようにするための項である．

2.5 シミュレーション環境

Fig.3で示したばね付き準受動歩行モデルにアクチュエータの出力をSACで学習させることにより，歩行と走行の実現を試みる．シミュレーションには物理エンジンMuJoCoを用いた．MuJoCoは接触の多い複雑な動的システムへの適用が可能であり，ロボット工学，バイオメカニクスなどの高速で正確なシミュレーションが必要とされる分野で使用される¹⁶⁾．式(2.2)を報酬関数として使用し， ω_2 の値のみを変更してシミュレーションを行った．学習は1000万ステップ行った．

3 シミュレーション結果

ω_2 を変えて学習させたところ， ω_2 が小さいときには歩行を，大きいときには走行を実現することに成功した．

3.1 歩行($\omega_2 = 0.025$)

Fig.5に歩行時のエネルギー変化の様子を示す．ここで，ピンク色の領域は右脚のみが支持脚となっている期間，青色の領域は左脚のみが支持脚となっている期間，紫色の領域は左右の脚がどちらも支持脚，つまり両脚支持期のときである．ピンク色または青色の領域から紫への領域へと変わるときに遊脚が地面に着く(例: 図中の濃い赤の丸)．シミュレーション開始から1~51[s]のモデルの総エネルギー

ギー(total energy)の平均値は294.5[J]，標準偏差は25.4[J]であった．Fig.6, Fig.7にはそれぞれモデルが受ける床反力，アクチュエータの出力を示す．上が右脚，下が左脚である．Fig.7ではアクチュエータの出力が正のとき，脚を床に押しつける力が働き，負のときは脚を持ち上げて床から離れる方向に力を生む．左右両方合わせたアクチュエータの仕事率は414.2[J/s]であった．

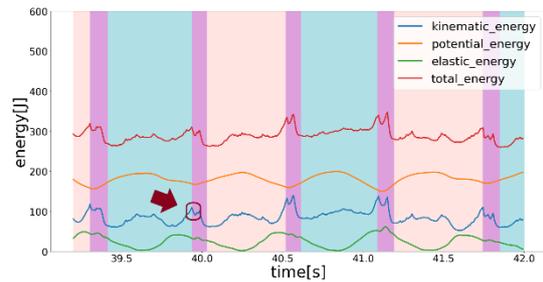


Fig. 5 歩行時のエネルギー変化

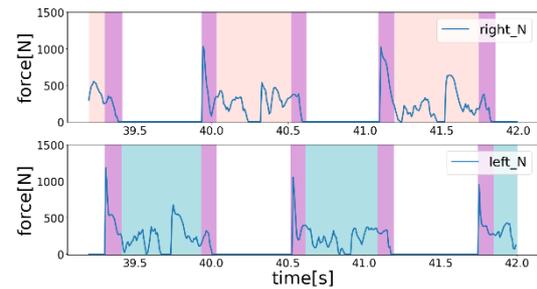


Fig. 6 歩行時の床反力

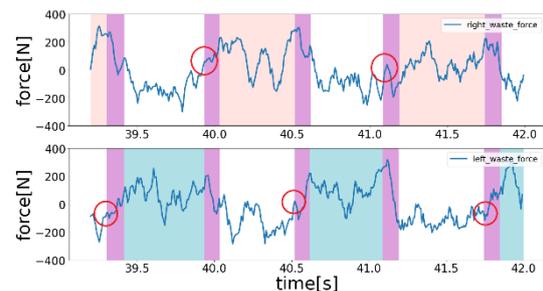


Fig. 7 歩行時のアクチュエータの出力

3.2 走行($\omega_2 = 1.2$)

走行時のエネルギー変化の様子をFig.8に示す．白色の領域は両脚とも地面についていない両脚遊脚期である．図中の濃い赤の丸で

示すように遊脚が地面に着くときに運動エネルギーが大きく減少していることがわかる。また、衝突直後に運動、位置エネルギーは減り、弾性エネルギーが増えていることから、衝突時に運動、位置エネルギーの一部は弾性エネルギーとして身体に蓄えられていることがわかる。シミュレーション開始から1~51[s]のモデルの総エネルギーの平均値は412.7[J]、標準偏差は28.2[J]であった。Fig.9、Fig.10にはモデルが受ける床反力、アクチュエータの出力を示す。左右両方合わせたアクチュエータの仕事率は547.7[J/s]であった。

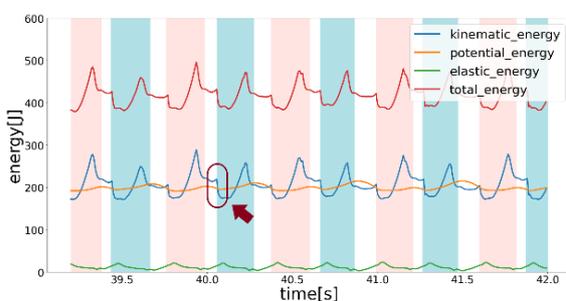


Fig. 8 走行時のエネルギー変化

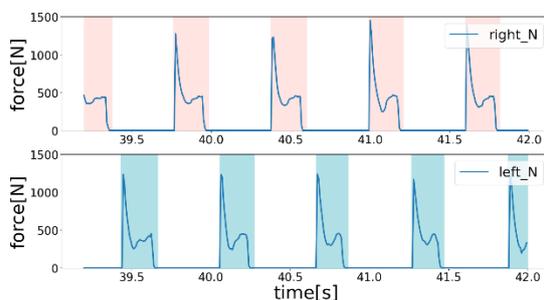


Fig. 9 走行時の床反力

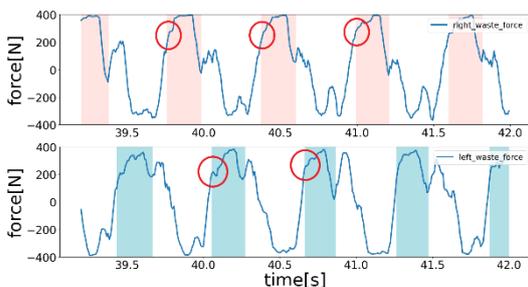


Fig. 10 走行時のアクチュエータの出力

4 考察

4.1 報酬関数とシミュレーション結果

式(2.2)のモデルの速度重み係数 ω_2 が小さいときには歩行、大きいときには走行の発現に成功した。よって、本研究で使用したばね付き準受動歩行モデルでは、速度の重視度を決め、エネルギー消費が少ない安定した運動を学習させることで歩行と走行がそれぞれ実現できると考えられる。

4.2 歩行と走行の違い

遊脚が支持脚となるときに注目した。Fig.7、Fig.10の赤丸からわかるように、歩行では脚が床に衝突するときにアクチュエータの出力がほぼ0であり静的な着地を行うのに対し、走行では出力が大きな正の値、つまり床に向かって脚を勢いよく打ちつけていることがわかる。その結果、Fig.5、Fig.8の濃い赤丸で示すように走行では歩行時に比べて衝突時に失う運動エネルギーが非常に多い。アクチュエータの仕事率は歩行時が414.2[J/s]、走行時は547.7[J/s]であり、歩行時の方が少ないエネルギーで移動していることから、歩行では静的な着地などを行うことでエネルギー損失を抑えているのではないかと考える。

一方、走行時は脚を勢いよく床に打ちつけている。ヒトが前方に推進するのは地面がヒトに及ぼした力の結果であるので¹⁷⁾、走行では大きな床反力を獲得することで、前に進むための推進力を得るために、このように脚を床に勢いよく打ち付けていると考えられる。Fig.6、Fig.9から、支持脚中の平均床反力を求めると、歩行では248.9[N]、走行では403.0[N]となり、たしかに走行時の方が大きな床反力を得ていることがわかる。

5 結言

本研究では、深層強化学習を用いてばね付き準受動歩行モデルにアクチュエータの使い方を学習させることで、準受動的な歩行および走行の実現に成功した。発現させたい歩容が歩行のときは速度の重視度を小さく、走行

のときは大きくし、エネルギー消費が少ない安定した移動を行わせるように報酬関数を設計すればよいことが示された。

また、歩行と走行の違いについても比較をした。歩行では脚を床に静かに衝突させることでエネルギー損失を少なくしており、逆に走行では脚を勢いよく打ちつけることで前に進むための大きな床反力を得ていると考えられる。

今後の課題としては、モデルのパラメータを変えてシミュレーションを行った場合、どのように歩容が変わるかの調査が挙げられる。また、今回は歩行と走行でそれぞれ別にシミュレーションを行ったが、歩行と走行の安定した切り替えを行い、歩行と走行を連続的に再現できるようにすることも挙げられる。

参考文献

- 1) Frederic J. Diedrich, Willian H. Warren, Jr.: Why Change Gaits? Dynamics of the Walk-Run Transition, Journal of Experimental Psychology: Human Perception and Performance 1995, vol.21, No.1, 183/202 (1995)
- 2) 衣笠哲也, 大須賀公一, 土師貴史: 受動歩行ロボットのすすめ, 5/29, コロナ社 (2016)
- 3) 土屋和雄: 身体適応-歩行運動の神経機構とシステムモデル, 148/153, 237/259, オーム社 (2010)
- 4) Steve Collins, Andy Ruina, Russ Tedrake, Martijn Wisse: Efficient Bipedal Robots Based on Passive-Dynamic Walkers, SCIENCE, vol. 307, No. 5712, 1082/1085 (2005)
- 5) Deniz Kerimoglu, Omer Morgul, Uluc Saranlı: Stability of a compass gait walking model with series elastic ankle actuation, 2015 International Conference on Advanced Robotics (ICAR), 351/356, (2015)
- 6) 清水寛子: 深層強化学習を用いたばね付き準受動歩行モデルにおける歩容生成, 修士論文 (2021)
- 7) R. McNeill Alexander: Walking and Running, American Scientist, Volume72, 348/354 (1996)
- 8) George H. Z. Liu, Michael Z. Q. Chen, Yonghua Chen: When joggers meet robots: the past, present, and future of research on humanoid robot, Bio-Design and Manufacturing 2, 108/118 (2019)
- 9) Andrew.A.Biewener: Patterns of Mechanical Energy Change in Tetrapod Gait: Pendula, Spring and work, journal of experimental zoology 305A:, 899/911 (2006)
- 10) Dai Owaki, Koichi Osuka, Akio Ishiguro: On the Embodiment That Enables Passive Dynamic Bipedal running, 2008 IEEE International Conference on Robotics and Automation, 341/346 (2008)
- 11) Z. Yue, J. Mester: A model analysis of internal loads, energetics, and effects of wobbling mass during the whole-body vibration, Journal of Biomechanics, 639/647 (2002)
- 12) Nikooyan AA, Zadpoor AA: Mass-spring-damper modelling of the human body to study running and hopping--an overview, Proc Inst Mech Eng H. doi: 10.1177/0954411911424210. PMID: 22320052, 1121/35 (2011)
- 13) 森北肇: 学習とそのアルゴリズム, 155/180, 森北出版 (2002)
- 14) Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Piter Abbeel, Sergey Levine: Soft Actor Critic Algorithm and Applications, Cornell University, (2018)
- 15) Whenhao Yu, Greg Turk, C.Karen Liu: Learning Symmetric and Low-Energy Locomotion, ACM Transactions on Graphics 37(4), (2018)
- 16) E. Todorov, T. Erez and Y. Tassa: MuJoCo: A physics engine for model-based control, 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5026/5033, (2012)
- 17) 金子公宥, 福永哲夫: バイオメカニクス—身体運動の科学的基礎, 8章, 杏林書院 (2004)