

## 進化的学習を通じたエージェントの立位保持に関する検討

山田 雅敏<sup>1</sup>, 高田 亮介<sup>2</sup>, 坂本 孝丈<sup>3</sup>, 竹内 勇剛<sup>4</sup>

### A Study on Standing of Agents through Evolutionary Learning

<sup>1</sup> YAMADA Masatoshi, <sup>2</sup> TAKATA Ryosuke, <sup>3</sup> SAKAMOTO Takafumi, <sup>4</sup> TAKEUCHI Yugo

#### 要 旨

本研究では, 身体的特徴が異なる3体のエージェントによる立位保持の獲得過程を分析し, 共通した学習プロセスを導くことを目的とする。学習手法として, 深層強化学習の中でも進化的学習 NEAT (Neuro-evolution of Augmenting Topologies) を採用した。また, エージェントへの報酬として, 重心位置が高いほど高い報酬を与えるように設定した。実験の結果, Humanoid-v2 は立位を保持できなかったのに対して, Hopper-v2 と BipedalWalker-v2 は立位保持を獲得したことが認められた。考察から, エージェントの上部の固定性と下部の動的性との連関により, 環境からの外力に対して対応可能となり, 立位を保持している可能性が示唆された。

**キーワード:** エージェント, 進化的学習, ニューロエボリューション, 立位保持

#### Abstract

The purpose of this study is to analyze processes of standing by three agents with different physical characteristics through evolutionary learning. The method called NEAT (Neuro-evolution of Augmenting Topologies) in deep reinforcement learning was adopted for three agents learning. The higher the center of gravity of the agent's body, the higher rewards for agents. As a result of this experiment, it was confirmed that Humanoid-v2 could not maintain standing, whereas Hopper-v2 and BipedalWalker-v2 acquired standing. From the discussion, the association between the upper fixation and the lower dynamics of the agent made it possible to respond to external forces from the environment, suggesting that the agent maintains standing.

**Keywords:** Agent, Evolutionary Learning, Neuro-evolution, Standing

---

<sup>1</sup> 常葉大学 Tokoha University

<sup>2</sup> 静岡大学大学院総合科学技術研究科 Graduate School of Integrated Science and Technology, Shizuoka University

<sup>3</sup> 静岡大学創造科学技術大学院 Graduate School of Science and Technology, Shizuoka University

<sup>4</sup> 静岡大学創造科学技術大学院 Graduate School of Science and Technology, Shizuoka University

## 1. はじめに

### 1.1. エージェントの動作獲得

人工知能の領域において、仮想空間上のエージェントが自ら動作を獲得する研究に関心が寄せられている (山田 2020)。たとえば、環境とのインタラクションを通して、エージェントが壁や裂け目などの障害物を自動で避ける動作を獲得したことが報告されているが (Heess, et al. 2017)、深層強化学習によりエージェントが自動で振る舞いを獲得することにより、エージェントの動作の学習プロセスが工学的に明らかになると期待される。

ここで、深層強化学習を用いてエージェントが自律的に動作を獲得するためには、エージェントに対する報酬設定が重要となる。一般的な設定方法として、(1) 目的の動作に沿った報酬を詳細に設定する、(2) 逆強化学習の手法によりデモンストレーションから報酬関数を導き出す、の2つのアプローチが考えられる。はじめに、(1) のアプローチは人間が報酬を詳細に設定するため、基本的に人間の想定内の学習プロセスと動作が実行されることとなる。また、人間が報酬関数を厳密に定義するためには、学習させたい動作に関する専門的な知識が必要となる課題も挙げられる。

次に(2)のアプローチは、深層強化学習によりエージェントが自らビッグデータを学習するが、そのプロセスはブラックボックスであり、ニューラルネットワークの結合荷重等に意味を見出すことは難しく、エージェントの行動原理が十分に説明できない。しかし一方で、人間が想定していなかった動作や学習プロセスを導き出せる可能性も考えられる。

### 1.2. 注目する動作と本研究の目的

本研究では、立位保持の動作に注目する。なお、エージェントが一定時間、立位を維持した状態を立位保持と定義する。そして、身体的特徴が異なる3体のエージェントによる立位保持の動作獲得の過程をトライアンギュレーション (方法論的複眼) により分析し、共通した学習プロセスを導くことを目的とする。これらの成果から、エージェントの動作獲得に関する学習プロセスの解明に寄与できると期待される。

## 2. 既往研究と理論的背景

人工知能に関する研究では、人間の知的な振る舞いを人工的に再現する試みが行われている。ある機能に注目し、その機能を再現し得る構造を明らかにすることで、人間の知的な振る舞いを理解しようとする。また、人工的なエージェントを構築することで、人間の知覚や行動、

認知プロセスを明らかにする方法がある。ここでいうエージェントとは、それ自体が目的を持ち、その目的に応じて自律的に行動するシステムを指す (Pfeifer, Bongard 2010)。エージェントは自律的に環境や他のエージェントなど外界と相互作用することで、エージェント自身が持つ目的を達成するように振る舞うこととなる。

エージェントを構築し振る舞いを再現するためには、知覚した情報と行動の間の対応関係を設計する必要があるが、このとき大きく2つのアプローチが考えられる。一つ目のアプローチは、既存の理論や設計者の経験に基づき知覚に対する行動をルールベースで記述する方法である。ルールベースで行動を記述する場合は、エージェントが知覚し得る環境の状態に対して、直接的にエージェントの行動が記述される。ルールベースで人と同じように振る舞うエージェントを設計する場合、実際の人の行動を分析し、行動のモデルを構築したうえで、人工的なエージェントによる再現が試みられる。この場合、エージェントを構築する目的はアプリケーションとしての実装である場合が多い。

また、エージェントによるシミュレーションやロボットの構築をモデル化手法の拡張として位置づけることもできる (Mergner 2010)。現象を抽象化して記述したモデルを再度具象化することで、モデルの妥当性の検証や修正を行うことができる。例えば、モデル化を行う際には注目されずモデルに組み込まれていなかった要素が、実際にはロボットを構築するうえで大きな障害になることがある。この時にモデルが扱う要素を拡張する必要性が明らかとなる。立位などの姿勢に関する研究については、数理的なモデルや計算モデルが提案されている。また、それらのモデルは実際にロボットに実装され、モデルの精緻化が行われている (Chiba, et al. 2016)。

次に、二つ目のアプローチでは、エージェントの構成が先行し、理論化や形式化は構成したエージェントを詳細に分析することを通して行われる。このようなアプローチは構成論的アプローチと呼ばれ (Pfeifer, Bongard 2010)、エージェントを構成するよりも前に注目する振る舞いをモデル化する必要はない。代わりに、エージェント自身が知覚と行動の間の関係性を環境とのインタラクションを通して学習することになる。ただし、エージェントが環境の状態を評価する関数 (適合度を求める関数や効用関数) を設計する必要がある。そして、何かしらの学習アルゴリズムによりエージェントが行動を獲得する場合、学習した知覚と行動の間の対応関係が分析の対象となり得る。どのような条件 (環境, 入力, 出力, 情報処理機構) において、エージェントが目的を達成する行動を獲得し得るのかを検証することができ、

目的を達成する行動が獲得される過程を分析することもできる。

また、構成論的アプローチでは新しくエージェントを構成するため、自然に存在する対象とは異なるパラメータや前提条件を設定することもできる。これにより既に構築されている理論やモデルを拡張し得る。構成論的アプローチによりエージェントに動作を獲得させる場合、強化学習や進化計算など機械学習の手法が用いられる。たとえば、Matsushita (2006) はエージェントの形態と運動制御を進化計算により学習させることで、二足歩行を行うエージェントを構成し、形態と運動制御の間の関係を分析している。また Ha (2019) は、強化学習により環境に適したエージェントの形態と行動を学習させる手法を提案している。一方で、エージェントの立位保持については、構成論的アプローチによる検証は管見の限り報告されていない。

そこで本研究では、構成論的アプローチに基づき仮想環境内で立位を保持するエージェントを構成し、獲得された行動を分析する。そこから、エージェントの形態の違いが、立位を保持する動作の獲得に与える影響を検証する。また、立位を保持するうえで、形態が異なるエージェント間に共通する行動について検討を行う。

### 3. 実験

#### 3.1. 立位保持の定義

Open AI Gym にて提供されているエージェントの初期状態から転倒せずに一定時間、立ち続けた状態を立位保持と定義する。

#### 3.2. 学習の手法

機械による学習の代表的な手法として、Q 学習が挙げられる (Watkins, Dayan 1992)。Q 学習の特徴として、価値関数を表形式で表現する点にあり、一つの状態に対して表 1 行が割り当てられることにより、全ての状態に対する行動価値が定義される。エージェントの学習プロセスはホワイトボックスである一方で、すべてのデータを離散化する必要があるため、本研究のような状態数が膨大となる連続的な動作に対する学習は、演算に多くの時間がかかり適さないと考えられる。次に、状態数が膨大である場合の学習として、Deep Q-Network (Volodymyr, et al. 2013) が挙げられる。この学習は、価値関数をニューラルネットワークにより近似する手法であるため、状態を離散化する必要はない一方で、局所解に陥る可能性が高いと考えられる。

そこで本研究では、NEAT (Neuro-evolution of Augmenting Topologies) に注目した (Stanley, Miikkulainen

2002)。NEAT は、ニューラルネットワークに遺伝的アルゴリズムを組み合わせたもので、ニューロン間の結合荷重とニューラルネットワークのトポロジーを遺伝子情報として進化させる特徴を持ち、生物学的進化との類似性が指摘されている。また、突然変異を発生させることにより、局所解に陥り難いという利点もある。さらに実装が容易で学習時間が短いという研究遂行上の利点から NEAT を採用した。

#### 3.3. 実験環境

本研究の実験環境は、表 1 の通りである。

表 1. 実験環境

学習手法 (設定)	NEAT 世代数：100 個体数：50 突然変異率：0.025
オペレーティングシステム	Windows10
プログラミング言語	Python 3.7.3
シミュレーション環境	OpenAI Gym NEAT-Python 0.92

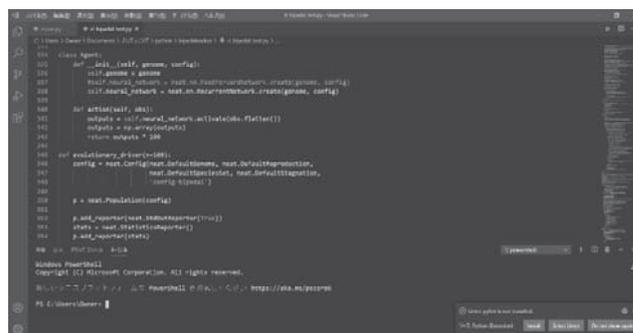


図 2 プログラムの実行画面 (Python)

#### 3.4. エージェントに対する報酬

三次元の二足エージェント Humanoid-v2 (図 1 左)、二次元の一足エージェント Hopper-v2 (図 1 中)、二次元の両足エージェント BipedalWalker-v2 (図 1 右) を構築した実験環境に用意し、立位を保持する学習をそれぞれ 5 回ずつ試行した。エージェントの重心位置が高いほど、エージェントに対して高い報酬を与えるように設定した。



図 1. 実験で使用した身体的特徴が異なるエージェント

## 4. 結果

### 4.1. 立位保持に関する定性的分析

Humanoid-v2 は初期状態から上半身を回旋させ、そのまま地面へ倒れ込み、立位を保持できなかったことが確認された(図3上参照)。次に、Hopper-v2は前後方向に振動を繰り返しながらも倒れずに、立位保持を獲得したことが認められた(図3中参照)。最後にBipedalWalker-v2は、片足を後方に振り上げた後、着地させ、立位保持を獲得したことが認められた(図3下参照)。

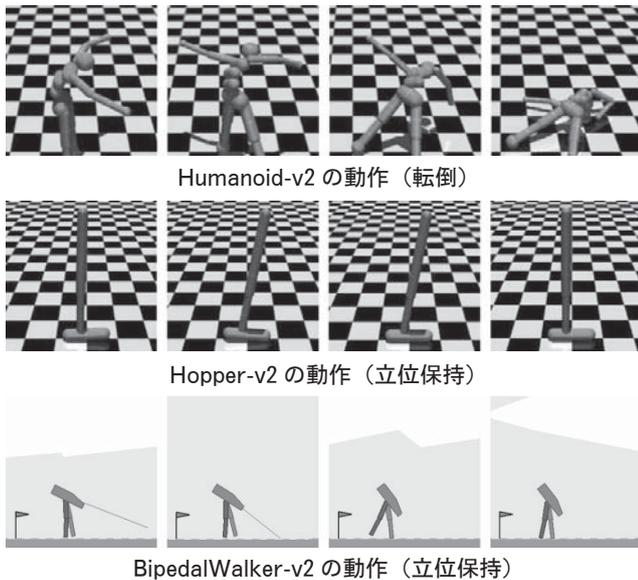


図3. 身体的特徴が異なるエージェントの動作結果

### 4.2. 立位保持を獲得したエージェントの定量的分析

立位保持を獲得した Hopper-v2 と BipedalWalker-v2 について、学習プロセスにおける入力値となるエージェントの身体部位の角度や速度などの情報を Excel CSV 形式により抽出し、定量的分析を行った。

はじめに、Hopper-v2 における上部関節の角度および上部関節の速度とともに、多少の振幅を繰り返しながらも収束したことが認められた(図4-1, 4-2 参照)。一方、下部関節の角度および下部関節の速度は、大きく振動していることが示された(図4-3, 4-4 参照)。

次に、BipedalWalker-v2 における上部船体部の角度は、最初に振動を繰り返しながら次第に収束し、一定の値に収束したことが認められた(図5-1 参照)。また上部船体部の速度は、多少の振幅を繰り返しながらも、角度と同じく収束する結果が示された(図5-2 参照)。一方、臀部の角度および膝の角度に関しては、周期性を有しながら振動していることが示された(図5-3, 5-4 参照)。

## 5. 結果

### 5.1. 上部の固定性

本研究では、ニューロエボリューションによる進化的学習を通して、身体的特徴が異なる3体のエージェントが立位保持を獲得する過程を分析し、共通した学習プロセスを導くことを目的とした。実験の結果、Humanoid-v2 は立位を保持できなかったのに対して、Hopper-v2 と BipedalWalker-v2 は立位を保持したことが認められた。

ここで、Humanoid-v2 の上部は頭頸部、肩部、腕部、体幹部と多くの部位で構成されており、上部の自由度は高く固定性は低いと推察される。また、Humanoid-v2 の立位は支持基底面に対して上方に重心を持つ逆振り子モデル(付録A 参照)として捉えることができるが、三次元空間上の Humanoid-v2 は上部を固定できずに回旋し、そのまま重心を保てずに倒れて立位が保持できなかったと推察される。

一方、Hopper-v2 の上部は関節が一つで構成される棒の形状であり、上部の自由度は低く固定性は高いと考えられる。また、BipedalWalker-v2 の上部の船体部は、加速・減速を繰り返しながらも角度を一定に保っていることから、同様に上部の固定性が高いと考えられる。

身体的特徴を見ると Humanoid-v2 に比べて、Hopper-v2 と BipedalWalker-v2 の上部関節は少ない。上部の固定性が高いエージェントが立位保持を獲得したことから、エージェントの立位保持の学習は、上部の固定性が関連していることが示唆された。

### 5.2. 下部の動的安定性

定量的分析の結果から、立位保持を獲得した Hopper-v2 と BipedalWalker-v2 のエージェントは、下部を揺らす動的状態であったことが認められた。ここで、BipedalWalker-v2 に関して、上部の船体部の数値が一定の値に収束したことが示されたことから、下部の振幅に関しては誤差ではないと考えられる。

エージェントが継続的に報酬を得るには、重心位置を上下に移動させるよりも、重心を高い状態に保ったまま静止する状態(静的状態)であることが有利であると考えられる。しかしながら、報酬が少なくなるにもかかわらず、立位を保持した2体エージェントは、下部の動的状態により重心位置を変化させたことが示された。たしかに、エージェントが静止している場合には、一定した報酬が与えられるわけであるが、環境からの外力に柔軟に対応することが難しい。一方で、エージェントが動的な状態である場合は、環境からの外力(例えば、地面からの反力)を受け流すことが可能となると考えられる。

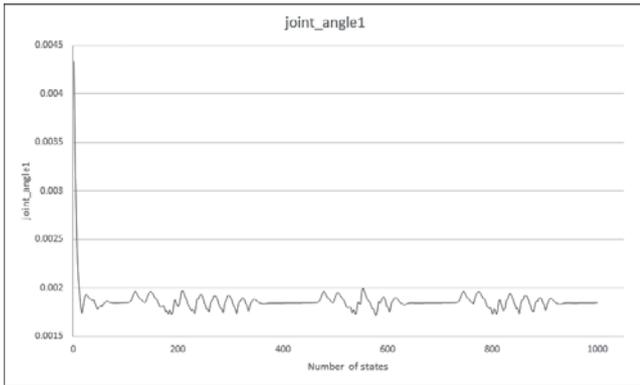


図 4-1. Hopper-v2 (上部関節の角度)

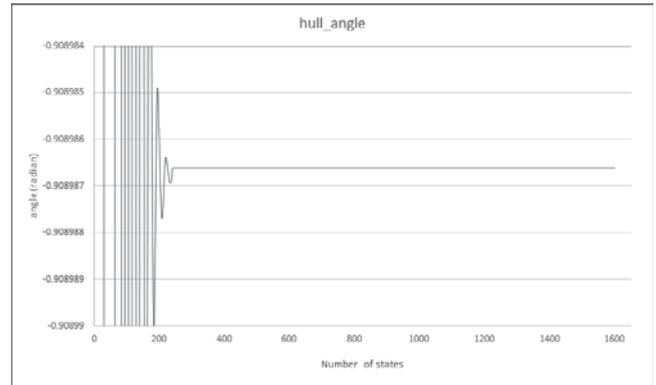


図 5-1. BipedalWalker-v2 (上部船体部の角度)

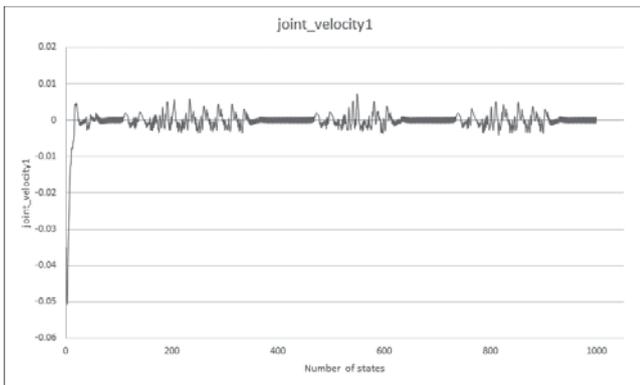


図 4-2. Hopper-v2 (上部関節の速度)

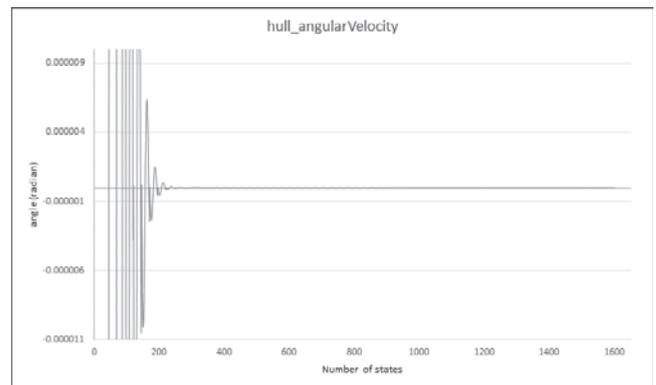


図 5-2. BipedalWalker-v2 (上部船体部の速度)

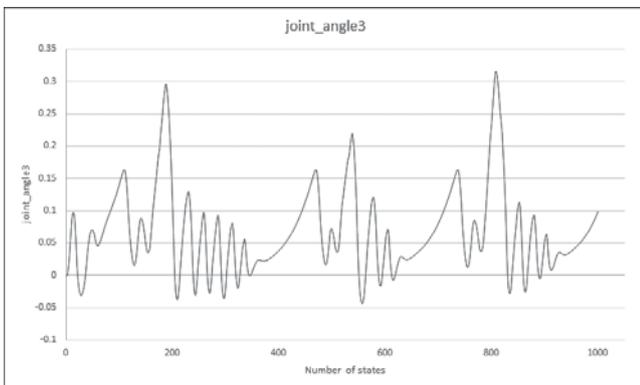


図 4-3. Hopper-v2 (下部関節の角度)

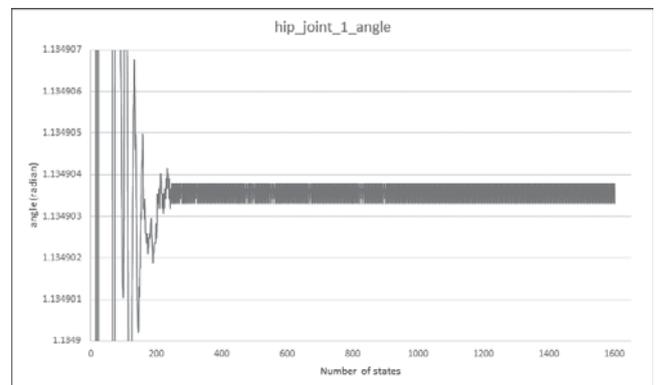


図 5-3. BipedalWalker-v2 (臀部の角度)

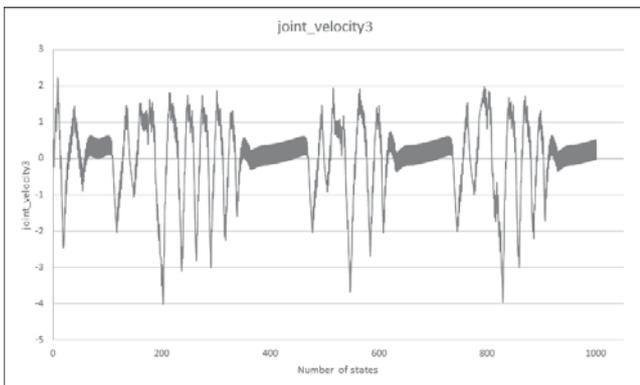


図 4-4. Hopper-v2 (下部関節の速度)

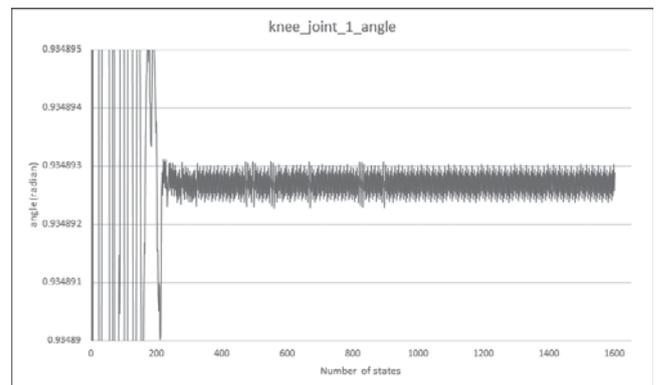


図 5-4. BipedalWalker-v2 (膝の角度)

図 4. Hopper-v2 の定量的分析の結果

図 5. BipedalWalker-v2 の定量的分析の結果

注

本研究の実験の結果から、環境に直接接している下部の動的性は、エージェントの立位保持の獲得に関連していることが示唆された。

このようにエージェントの上部の固定性と下部の動的性との連関により、環境からの外力に対して対応可能となり、立位を保持していることが示唆された。以上の考察に関して、エージェントの立位保持に関するモデルを図6に示す。

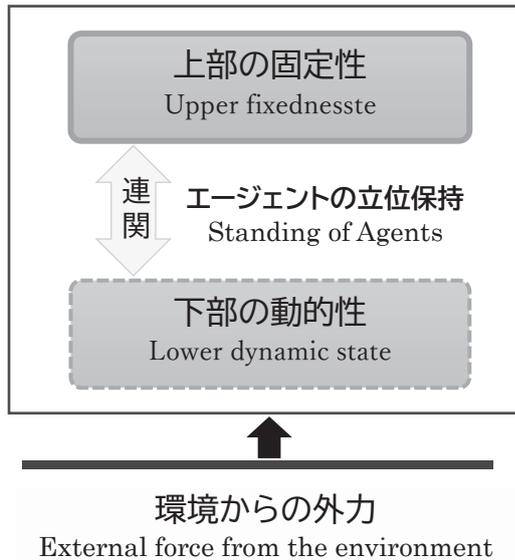


図6 エージェントの立位保持に関するモデル

6. まとめと今後の課題

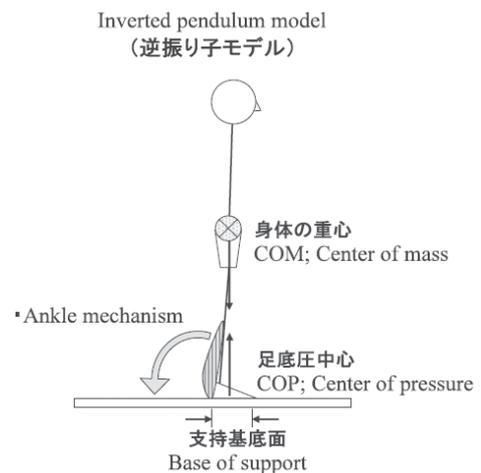
本研究では、深層強化学習の中でも進化的学習 NEAT を通して、身体的特徴が異なる3体のエージェント立位保持の過程を分析し、共通した学習プロセスを導くことを目的とした。実験の結果、Humanoid-v2は立位が保持できなかったのに対して、Hopper-v2とBipedalWalker-v2は立位保持を獲得したことが認められた。考察から、エージェントの立位保持の獲得に関して、エージェントの上部の固定性と下部の動的性が関連していることが示唆された。エージェントの立位保持に関する構成論的アプローチによる検証は管見の限り報告されていないことから、本研究の成果は一定程度の新規性を有し、当該領域の発展に寄与する知見が得られたと考えられる。

今後の課題として、得られた知見をより確実なものにするために、さらに異なる身体的特徴を持つエージェントに対しても同様の実験を行うことを視野に入れている。

- 1) 2020年度人工知能学会全国大会(第34回)で発表した研究の発展の位置づけとして、第1章, 2章, 5章を中心に加筆・修正を行った(25%以上)。また、実験結果の信頼性を高める目的で、3体のエージェントの学習をそれぞれ5回ずつ試行を行った。
- 2) Open AI Gymについては、次のURLを参照されたい。 <https://gym.openai.com/>
- 3) NEATについては、次のURLを参照されたい。 <https://neat-python.readthedocs.io/en/latest/>

A 付録

図Aの逆振り子モデルは、長谷公隆(2006)立位姿勢の制御;リハビリテーション医学, 43, 2006, pp.542-553より引用。



図A: 逆振り子モデル

参考文献

- [1] Chiba Ryosuke, Takakusaki Kaoru, Ota Jun, Yozu Arito, Haga Nobuhiko, "Human upright posture control models based on multisensory inputs: in fast and slow dynamics", Neuroscience research, 104, 2016, pp.96-104.
- [2] Ha, D., "Reinforcement learning for improving agent design", Artificial life, 25(4), 2019, pp.352-365.
- [3] 長谷公隆, "立位姿勢の制御;リハビリテーション医学", 43, 2006, pp.542-553.
- [4] Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, Martin A.,

- Silver, D., “Emergence of Locomotion Behaviours in Rich Environments”, arXiv preprint arXiv:1707.02286, 2017.
- [ 5 ] Matsushita Kojiro, Yokoi Hiroshi, Arai Tamio, “Pseudo-passive dynamic walkers designed by coupled evolution of the controller and morphology” , Robotics and Autonomous Systems, 54(8), 2006, pp.674-685.
- [ 6 ] Mergner, T., “A neurological view on reactive human stance control”, Annual Reviews in Control, 34(2), 2010, pp.177-198.
- [ 7 ] Mnih, V., Kavukcuoglu, K, Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M., “Playing Atari with Deep Reinforcement Learning”, arXiv preprint arXiv:1312.5602, 2013.
- [ 8 ] Pfeifer, R., Bongard, J., 細田耕 (訳), 石黒章夫 (訳), “知能の原理 - 身体性に基づく構成論的アプローチ”, 共立出版, 2010.
- [ 9 ] Stanley, Kenneth O., Miikkulainen R., “Evolving Neural Networks through Augmenting Topologies”, Evolutionary Computation, 10(2), 2002, pp.99-127.
- [10] Watkins, Christopher JCH., Dayan, P., “Q-learning” , Machine learning, 8(3-4), 1992, pp.279-292.
- [11] 山田雅敏, 高田亮介, 坂本孝丈, 竹内勇剛, “身体的特徴が異なるエージェントにおける立位熟達過程の分析 ニューロエボリューションによる進化的学習に注目して”, 人工知能学会, 2020年度人工知能学会全国大会論文集 (4Rin1-19), 2020.

