氏名：周佑綸
出身大学：国立台湾大学

研究テーマ：

強化学習を用いた四足歩行ロボットにおける環境適応の実現

研究の目的：

本研究は、四足歩行ロボットの環境適応能力を効果的に向上させる目的を持ち、強化学習（Reinforcement Learning, RL）を用いた手法を提案することである。RLとは、人工知能やコンピューターなどの学習者にデータを与え、環境との相互作用を通じて最適な行動を学ばせる「機械学習」手法の一つである[1]。この方法は、ロボットが複雑な環境で適切な行動を自律的に学習するための力強い手段として注目されている。

まず、RLを利用した四足歩行ロボットに関する先行研究として、マサチューセッツ工科大学（MIT）の研究[2]が挙げられる。この研究の関連記事[3]によると、従来のロボットの動きをコントロールする制御装置は解析的に設計され、様々な状況でロボットの反応をプログラムするため、人間の介入が必要である。このプロセスには非常に多くの時間がかかると指摘されている。ゆえに、本研究はこれまでの制御装置に関わる問題を解決し、四足歩行ロボットが困難な地形を効率的に乗り越えることを目指している。このような目標を持つ本研究は、ロボテックス分野において新たな可能性を見つけ出す役割を果たすことが期待される。

以下は提案の理由について詳しく説明する。初めに、あらゆる場合に対してロボットをプログラムする難しさと単調さにある。特定の地形でロボットが順調に進まない時に、従来の方法ではエンジニアが原因を調べ、制御装置を調整する必要がある。この過程は数々の労働力と時間を要する。しかし、提案手法の試行錯誤を用いれば、人間による定義された行動の必要性はなくなる。その結果、時間を節約し、より価値がある研究に専念できるようになる。

また、最近の研究[4, 5, 6, 7, 8, 9]では、ロボットが様々な地形でRLを用いた移動制御装置を学習させ、経験を通じて自動的に適応能力を向上できることが示された。さらに、現在のシミュレーションソフトウェアで、ロボットは短時間で多様な地形でのデータを積み、効率的な学習と改善を実現できる

ことが確認されている。以下では、先行研究[4, 5]の貢献について述べたいと考えられる。

チューリッヒ工科大学（ETH）の先行研究[4]において、困難な地形での「ブラインドな」四足歩行のための制御装置が提案された。ここでの「ブラインド」とは、カメラやライダー（LiDAR）などの外部センサーを使わずに、エンコーダと慣性計測装置のみを利用し、固有受容性を測定することである。なぜなら、外部センサーが地面の摩擦や変形などの物理特性を正確に測定できず[10]、また植生や雪、水などの障害物によって妨げられる可能性があると指摘された。

前述のような複雑な地面でもたくましい歩行制御装置を学習するため、この研究はカリキュラム学習を応用した。カリキュラム学習とは、学習する事例の順序を工夫し、簡単なものから学ぶ方法で、所要時間の短縮やモデルの高精度化手法である[11]。具体的にこの研究では、環境パラメータの分布を徐々に変更し、粒子フィルタリングを用いて中程度の難易度の地形パラメータの分布を維持しつつ[12, 13]、カリキュラム学習を実現する。この分布は学習に応じて適応することである。これにより、泥や瓦礫、草、雪、都市などの様々な環境での歩行を実現し、ロボットの移動能力を向上させ、未知の環境でも十分に通用することと繋がった。

最後に、ロボットの環境適応性に関する先行研究として、カーネギーメロン大学（CMU）とカリフォルニア大学バークレー校（UCB）の共同研究[5]が挙げられる。この研究では、参考文献に示されたRL技術と新しい適応モジュールを組み合わせた「素早いモーター適応」あるいは「Rapid Motor Adaptation」というアルゴリズムが提出された。この方法によって、ロボットは実際の状況をリアルタイムで把握し、瞬時に対応することができるようになる。

また、このアルゴリズムは異なる環境いおいて、事前に定義された動作モデルは必要なく、すなわち人間がプログラムする必要がなく、ロボットは固有受容データに基づいて階下を下りたり、岩の上を歩いたりする能力を手に入れる。この研究は、新たな手法によってロボットの適応性を向上させる非常に重要な貢献となっている。

まとめると、RLを利用することによる利点は以下のようになる。

1. ロボットの適用範囲は室内に限定されず、外の多様な状況においてもうまく対応できる能力を高める。

⇒ロボットの実用性が拡大し、実世界での応用が大幅に広がる。

（２）異なる地形でのプログラミングにかかる時間が短縮され、研究者はより高度な研究に集中することが可能になる。

さらに、

（３）RLを用いた方法は四足歩行ロボットの分野に限らず、他のロボットにも応用可能である。

⇒多様なロボットの制御や適応性の向上においても頼もしい解法となる可能性がある。

以上の三つのメリットが、強化学習の採用がロボット工学分野にもたらす価値を示している

よって、本研究は現実の状況に合わせて適応するための解決策を検討し、RLと適応アルゴリズムの組み合わせによって、四足歩行ロボット工学分野の進歩を促そうと考えている。さらに、この研究は、ロボットがシミュレーションを通じて自律的に学習し、その成果を現実世界での応用を可能にすることを求めている。これにより、四足歩行ロボットが短時間で環境適応の能力を高め、複雑な地形を通るときにも効果的に対応できるような開発に寄与することを目指している。本研究は、将来的なロボティク分野の発展に向けて重要な技術となることを目標としている。

研究の方法：

本研究では、四足歩行ロボットの環境適応能力向上のために、
強化学習（Reinforcement Learning, RL）を活用する手法を提案し、その実現に向けて詳細な研究を行うことを目指している。具体的な研究方法と手順は以下のようになる。

1. 文献調査と先行研究の分析：
まず、四足歩行ロボットの環境適応をめぐっての先行研究を調査し、論文を集める。例えば、MITやETH、CMUなどの大学や関連する研究機関における四足歩行ロボットについての先行研究を参考し、RLを用いて環境適応能力を向上させるための成功事例や課題を深く理解する。最後に、これらの先行研究の成果と問題点を比較・分析しながら、新たな方法や改良点を見つけ出せるであろう。

2. 強化学習アルゴリズムを選ぶ：
ここでは、相応しい強化学習アルゴリズムを選定する。例えば、Q学習系（DQN、Gorila）、SARSA法、Actor-Critic系[14]などのアルゴリズムを検討す

る。これらのアルゴリズムについて深く理解し、それぞれの特性や適用範囲を評価する。そして、研究の目的に合わせて最適なアルゴリズムを選び出す。

3. シミュレーション環境の構築：
提案手法の評価を行うために、MITの研究[2]で使用されたIsaacGymシミュレータ[15]と、[8]のオープンソースリポジトリから適応したコードを用いると考えている。すなわち、公開されたプログラムのデータベースを利用する。IsaacGymで、現実世界の環境変化を再現し、ロボットの運動制御をシミュレーションする。

4. 強化学習の実装とモデル訓練：
四足歩行ロボットの動作制御を学習させるために、予て選択した強化学習アルゴリズムの実装を行う。そして、IsaacGymを用いて訓練を行い、ロボットが異なる環境で適切な行動を学習するプロセスを進める。訓練の過程では、アルゴリズムのハイパーパラメータチューニング（調整）[16]も行う。ここの「ハイパーパラメータ」とは、機械学習を行う前に予め人の手で設定する変数を指す。

5. モデル・実機実験の評価と改善：
学習が進むにつれて、シミュレーションでの結果を定期的に評価・改良する。そして、シミュレーションが一定のレベルに到達したら、実際の四足歩行ロボットに対して実機実験を行う。シミュレーション環境や実世界におけるロボットの表現を通じて、適応能力や動作の改善が必要なところを特定し、シミュレーションと実世界の違いも明確に把握する。最後に、アルゴリズムの有用性を評価し、アルゴリズムやモデルを適切な調整や改善の方向性を検討することを考えている。

6. 評価と結果の分析：
上記のステップを繰り返すことで、実験結果を分析し、RLアルゴリズムの性能と有効性を評価する。具体的には、IsaacGymにおけるシミュレーションの結果と実機実験の結果を比較し、ロボットの環境適応能力向上度や安定性を定量的に評価する。これにより、提案手法が実世界での適応性をどの程度改善するかを明確に把握し、改善の方向性を探求することを目指している。

7. 結論と展望：
評価結果に基づいて、本研究の有効性および限界について結論付ける。そして、今後の展望や改善点を議論し、提案手法を実世界での実装にどのように

応用できるかを検討する。さらに、この方法が四足歩行ロボットの分野にとどまらず、ロボティクス分野全体にもよい影響を及ぼす手段にも探し求める。

以上のステップに応じて、RLを応用する提案手法では、四足歩行ロボットの環境適応能力を効果的に向上させる方法を詳細に検討し、その有効性を確認することを目標としている。本研究によって得られる成果は、四足歩行ロボットの実用性や応用範囲が拡大し、ロボティクス技術の進展と未知の環境への適応能力を持つロボットの実用性向上に新たな可能性をもたらすことが期待される。

参考文献一覧：

[1] https://www.brainpad.co.jp/doors/news_trend/about_reinforcement_learning/
[2] G. B. Margolis, G. Yang, K. Paigwar, T. Chen, and P. Agrawal. Rapid locomotion via reinforcement learning. Robotics: Science and Systems, 2022. doi:10.48550/arXiv.2205.02824.
[3] https://news.mit.edu/2022/3-questions-how-mit-mini-cheetah-learns-run-fast-0317
[4] Joonho Lee, Jemin Hwangbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning quadrupedal locomotion over challenging terrain. Sci. Robot., 5 (47): eabc5986, October 2020. doi:10.1126/scirobotics.abc5986.
[5] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid motor adaptation for legged robots. In Proc. Robot.: Sci. and Syst. (RSS), Virtual, July 2021. doi:10.48550/arXiv.2107.04034.
[6] Gabriel B Margolis, Tao Chen, Kartik Paigwar, Xiang Fu, Donghyun Kim, Sangbae Kim, and Pulkit Agrawal. Learning to jump from pixels. In Proc. Conf. Robot Learn. (CoRL), pages 1025–1034, London, UK, November 2021. doi:10.48550/arXiv.2110.15344.
[7] Takahiro Miki, Joonho Lee, Jemin Hwanbo, Lorenz Wellhausen, Vladlen Koltun, and Marco Hutter. Learning robust perceptive locomotion for quadrupedal robots in the wild. Sci. Robot., 7(62): abk2822, January 2022. doi:10.1126/scirobotics.abk2822.
[8] Nikita Rudin, David Hoeller, Philipp Reist, and Marco Hutter. Learning to walk in minutes using massively parallel deep reinforcement learning. In Proc. Conf. Robot Learn. (CoRL), pages 91–100, London, UK, November 2021. doi:10.48550/arXiv.2109.11978.
[9] Jonah Siekmann, Kevin Green, John Warila, Alan Fern, and Jonathan Hurst. Blind bipedal stair traversal via sim-to-real reinforcement learning. In Proc. Robot.: Sci. and Syst. (RSS), Virtual, July 2021. doi:10.48550/arXiv.2105. 08328.

[10] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, R. Siegwart, Robot-centric elevation mapping with uncertainty estimates, in Mobile Service Robotics (World Scientific, 2014), pp. 433–440. doi:10.1142/9789814623353_0051
[11] https://confit.atlas.jp/guide/event-img/jsai2018/3Pin1-09/public/pdf?type=in
[12] J. C. Brant, K. O. Stanley, Minimal criterion coevolution: A new approach to open-ended search, in Genetic and Evolutionary Computation Conference (GECCO, 2017), pp. 67–74. doi:10.1145/3071178.3071186
[13] R. Wang, J. Lehman, J. Clune, K. O. Stanley, Paired open-ended trailblazer (POET): Endlessly generating increasingly complex and diverse learning environments and their solutions. doi:10.48550/arXiv.1901.01753
[14] https://qiita.com/shionhonda/items/ec05aade07b5bea78081
[15] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac Gym: High performance GPU-based physics simulation for robot learning. arXiv preprint, 2021. doi:10.48550/arXiv.2108.10470.
[16] https://www.codexa.net/hyperparameter-tuning-python/