計測自動制御学会東北支部第336回研究集会(2021.12.1)
資料番号336-5

深層強化学習を用いた足場配置計画による６脚ロボットの歩行動作獲得

Walking Motion Acquisition of Hexapod Robot by Foothold Placement Planning Using Deep Reinforcement Learning

○中川原拓海*，釜谷博行*，原元司**，工藤憲昌*
Takumi Nakagawara*, Hiroyuki Kamaya*, Motoshi Hara**, Norimasa Kudoh*
*八戸工業高等専門学校，**松江工業高等専門学校
*National Institute of Technology, Hachinohe College,
**National Institute of Technology, Matsue College

キーワード: 深層強化学習(Deep Reinforcement Learning), 6脚ロボット(Hexapod Robot),
足場配置計画(Foothold Placement Planning), 不整地(Rough Terrain)
連絡先: 〒039-1192 青森県八戸市田面木字上野平16-1 八戸工業高等専門学校　産業システム工学専攻
Tel.: 0178-27-7283，E-mail: kamaya-e@hachinohe-ct.ac.jp

1. はじめに

近年，多脚ロボットは惑星探査や救助活動の分野において活躍が期待されている．多脚ロボットは脚の自由度が高いため，不整地に対して体の水平を保つことができる，障害物を乗り越えることができるなど，車輪型ロボットでは移動できないような地形でも行動することができる．しかし，自由度が高いため，制御が複雑になるという問題がある．

この問題を解決するアプローチとして，強化学習(Reinforcement Learning；RL)を用いてセンサ情報からモータ制御信号を直接計算する方法がある．しかし，得られる方策は，ロバスト性が低く [1]，複雑な報酬関数の設計や大量の学習サンプルが必要となる [2]．また，学習した方策はわずかな地形の変化でも失敗することがある．

そこで本研究では，地形の変化に対応するために，足場の配置計画のための強化学習と脚の追従制御に分けて多脚ロボットの制御を行う新たな手法を提案する．モデルフリーのRLアルゴリズムであるSoft Actor Critic（SAC）[3]を用いて，ロボットの状態と地形の情報をもとに望ましい足の位置を計算する方策を学習する．本発表では，不整地環境での前進動作の獲得を目的とする．

2. 開発環境

本研究では，Trossen Robotics 社の PhantomX [4]を用いた(Fig.1)．このロボットは，3 つの関節を制御する 6 本の脚で構成されている．各脚の 3 個のモータのうち，第一関節は脚を前後に動かし，他の 2 つの関節は脚をそれぞれ持ち上げるように動作する。

実験環境として，ロボットを制御するために Robotic Operating System（ROS）を使用する．また，ROS のサポートが充実していることから，シミュレーション環境 Gazebo を使用する．使用するセンサとアクチュエータ

のデータの時間を揃えるために，25Hz の固定間隔でデータの取得を行う．また，パーリンノイズを用いて，Fig.2 に示すような標高マップを作成し，地形の自動生成を行う．

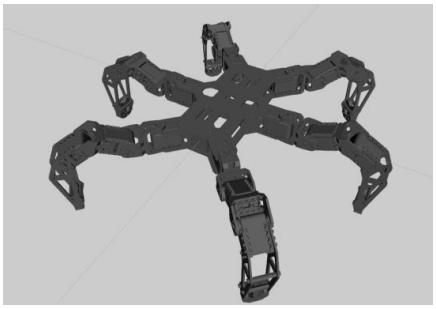

Fig.1 ロボットモデル

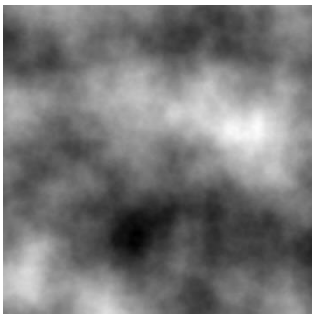

Fig.2 生成した標高マップ

3. 強化学習

強化学習は，ある環境下においてエージェント(学習主体)が試行錯誤を通じて得られる報酬をもとに，より良い行動を学習によって獲得する手法である．エージェントと環境は以下のやり取りを行う．

(1) エージェントは時刻tにおいて環境の状態$s(t)$に応じて，行動$a(t)$を出力する．
(2) エージェントの行動を実行し，環境は$s(t+1)$へ状態遷移し，その遷移に応じた報酬$r(t)$をエージェントに与える．
(3) 時刻をtから$t+1$へと進め，(1)~(3)のサイクルを繰り返す．

エージェントは累計報酬の最大化を目的とし，状態から行動を出力する方策$\pi(s)$を獲得する．

本研究では，連続的な制御タスクを扱っているため，状態空間は連続的であり，モータ制御に対応する行動空間も連続的である．そこで，モデルフリー学習で特に連続値制御のタスクにおいて広く用いられている SAC を使用した．SAC では報酬の最大化に加え，方策のエントロピーを考慮した(1)式のような目的関数を最大化する方策πを求める．

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}_{\pi}[r_t + \alpha \mathcal{H}(\pi)] \qquad (1)$$

エントロピー$\mathcal{H}(\pi)$を考慮することで行動の多様性を保ちながら，かつ，得られる報酬を最大化することができる．αは重み係数である．また，オフポリシーで学習することからサンプル効率に優れており，学習率等のハイパーパラメータの変動に対して頑健に動作する．

4. 手法

提案システムの全体図を Fig.3 に示す．

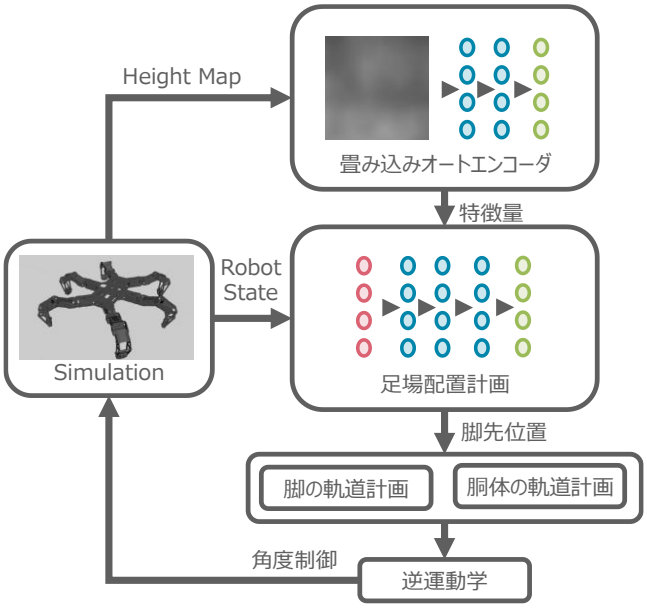

Fig.3 提案システム概要

システムの動作は大きく分けて，以下のようになる．

(1) 畳み込みオートエンコーダによるロボット周辺地形の特徴量抽出

(2) 特徴量とロボットの状態から足の接地位置を出力
(3) 脚先位置に追従するための脚と胴体の軌道生成
(4) 逆運動学により角度制御命令に変換

4.1 地形情報の取得

地形を考慮した歩行を行うために，ロボット周辺の標高マップ（Height Map）の特徴量を足場配置計画の状態に導入する．特徴量の抽出には教師なし学習の 1 つである畳み込みオートエンコーダ [5]を用いる．ここでは，ロボットを中心とする 64×64 次元の標高マップを用いる．これにより，標高マップから特徴を抽出し, 64 次元のベクトルとして表現する．

4.2 足の接地位置の計算

ここでは，次に動かす足の接地位置を求める．次に動かす脚は，同側の前脚・後脚と反対側の中脚がほぼ同じタイミングで地面に接地するトライポッド（tripod）歩容をベースにして決定する．

行動空間は，動かす 3 つの足のx座標とy座標を用いた．よって計 6 次元の行動空間となる．z座標は，出力された足の x 座標とy座標の位置に対する標高マップから直接計算する．

状態空間は各足の位置 3×6 次元，ロボットの姿勢（ロール角，ピッチ角，ヨー角）の 3 次元，標高マップの特徴量 64 次元を用いた．よって計 85 次元の状態空間となる．

報酬関数rは安定した前進動作を獲得するために(2)式のように設定し，学習を行った．各時間 step において報酬関数によって報酬が得られる．

$$r = v + SM - \sum_{i}^{6} \left\| \hat{f}_i - f_i \right\|^2 \qquad (2)$$

ここで，vは胴体の速度，$\hat{f}_i$は方策から出力された足位置, f_iは実際の足位置を表す. また，SMは安定余裕であり，ロボットの重心位置(Center of Mass ; COM）の地面への投影点から支持脚によって形成される多角形までの最短距離である．この指標が大きいほどロボットは安定して歩行しているといえる．この報酬関数によって安定した歩行のための足位置を出力する方策を学習することができると期待できる．

学習には，3 つの隠れ層(256 個の隠れユニットを持つ)を持つニューラルネットワークを使用した。

4.3 脚と胴体の軌道計算

胴体の最終位置は，目標足位置で形成される三角形において安定余裕が最大になる位置とした．ロボット胴体の動きは完全に支持脚に依存するため，胴体の動きをロボットの各支持脚の関節空間に変換する必要がある．この目的のために, 胴体から N 個の経路点を収集し，運動学的な計算によって，各支持脚の関節の経路点を得る．次に，これらの経路点を三次スプライン曲線で補間し，滑らかな関節軌道を得る．θ_1^j，θ_2^j $\cdots$ θ_N^jを各支持脚関節の対応する経路点とする. ここで, $j = 1,2,3$はある支持脚の 3 つの関節，Nは経路点の数である．支持脚の任意の関節jに対して，経路点と連続性の条件を満たすように，三次スプライン曲線Sを構築する．

$$\begin{cases} S(t_i) = \theta_i^j & (i = 1,2,\cdots N) \\ \lim\limits_{t \to t_i} S(t) = S(t_i) & (i = 1,2,\cdots N-1) \\ \lim\limits_{t \to t_i} S'(t) = S'(t_i) & (i = 1,2,\cdots N-1) \\ \lim\limits_{t \to t_i} S''(t) = S''(t_i) & (i = 1,2,\cdots N-1) \end{cases} \qquad (3)$$

また，この方程式を解くために，さらに境界条件を設定する．

$$\begin{cases} S'(t_1) = 0 \\ S'(t_N) = 0 \end{cases} \qquad (4)$$

これにより支持脚の軌道を得ることができる．

また，遊脚の経路は，障害物を避けて目的の足位置まで脚を運ぶために高速に探索が可能である RRT(Rapidly exploring random tree) [6]を用いて作成する．支持脚と同様に，これらの経路点を三次スプライン曲線で補間し，滑らかな脚軌道を得る．求めた脚軌道を逆運動学によって，各関節の角度制御命令に変換する．

5. 実験

5.1 前進動作の獲得

提案手法を用いて前進動作の獲得実験を異なるランダムシードを用いて15回行った．1000 step を 1 episode とし，合計 5000 episode の学習を行った．最初は平坦な地形で学習させ，1000 episode 以降は自動的に作成した標高マップによる不整地環境で学習を行った．不整地環境は徐々に標高を高くし，最終的に 0.10 m の標高マップで学習を行った．

結果として，前進動作を獲得することができた．報酬の平均値，最小値および最大値を Fig.4 に示す．横軸は episode 数を示す。グラフの実線は 15 回の平均値，塗りつぶされた部分は最小値と最大値の範囲を表している．報酬は約 2500 episode で収束し，学習には約 13 時間かかった．

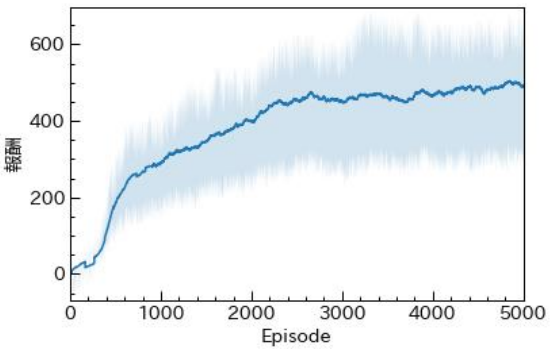

Fig.4 学習曲線

5.2 獲得した歩行動作の評価

学習時には使用しなかった新たな不整地で評価実験を行った．評価方法は，学習した方策を用いて，再度凹凸のある地形での報酬を計測した．最大高さが 0.00m，0.05m，0.10m，0.15 m の地形で，それぞれ 100 回のシミュレーション実験を行った．

実験結果を Table 1 と Fig.5 に示す．グラフの横軸は地形の最大高さ、縦軸は報酬の平均値および標準偏差を表す。結果として，学習時とほとんど変わらない報酬を得ることに成功した．しかし，0.0m の平坦な地形では，平均報酬が 412.5 と一番小さく，標準偏差が 54.2 と一番小さかった．一番単純な地形で報酬が一番小さくなった原因は，学習後半で平坦な地形を使わなかったことにあると考える．そこで，学習時に高さを徐々に大きくするのではなく，ランダムで高さを変更して学習することで平坦な地形でも高い報酬が得られる歩行が獲得できるものと考える．

また，0.05 m の地形では平均報酬が 572.7 と一番高くなり，標準偏差も 213.1 と一番大きくなった．以降，最大高さを上げると報酬が徐々に下がり，標準偏差も下がる結果となった．このような結果になった要因は，移動速度と安定余裕のトレードオフな関係にあると考える．平均移動速度に関しては，0.05 m の地形では 0.253 m/s，0.10 m の地形では 0.231 m/s，0.15 m の地形では 0.202 m/s となった．安定余裕に関しては，0.05 m の地形ではばらつきが多く，0.10 m，0.15 m の地形では安定余裕がほぼ一定だった．このことから 0.05 m の地形では，安定余裕が小さくても転倒する可能性が低いため，速度重視の歩行が得られたことが分かる．一方，最大高さが大きくなると転倒する恐れがあるため，安定余裕を重視することになり，移動速度が小

さくなることが分かる．このことから，地形の特徴を正確に把握して歩行を行っているといえる．

Table 1 地形最大高さに対する性能分布の比較

最大高さ[m]	平均値	標準偏差
0.00	412.5	54.2
0.05	572.7	213.1
0.10	502.3	159.8
0.15	468.2	121.9

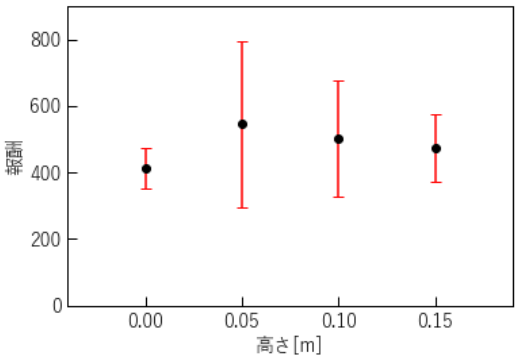

Fig.5 地形最大高さに対する性能分布の比較

6. おわりに

本研究では，深層強化学習を用いた足場配置計画を提案し，前進動作の獲得実験を行った．結果として，不整地でも安定した歩行を獲得することに成功した．また，学習に利用していない未知環境でも，学習で使用した地形と同等の歩行性能を示した．

今後の課題として，全方向への歩行動作の獲得やパーリンノイズ以外の様々な環境で学習により歩行動作を獲得することなどが挙げられる．

7. 謝辞

研究の一部は JSPS 科研費 JP19K03043 の助成を受けたものである．

参考文献

1) Peng, X. B., Coumans, E., Zhang, T., Lee, TW., Tan, J. and Levine, S., "Learning agile robotic locomotion skills by imitating animals", arXiv preprint arXiv:2004.00784, 2020.
2) Hwangbo, J., Lee, J., Dosovitskiy, A., Bellicoso, D., Tsounis, V., Koltun, V. and Hutter, M., "Learning agile and dynamic motor skills for legged robots", Science Robotics, 4(26), 2019.
3) Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P. and Levine, S., "Soft actor-critic algorithms and applications", arXiv preprint arXiv:1812.05905, 2018.
4) Trossen Robotics, “PhantomX AX Hexapod Mark Ⅱ Kit”, https://www.trossenrobotics.com/hex-mk2 (2021 年 11 月 21 日閲覧)
5) Lovedeep Gondara, "Medical image denoising using convolutional denoising autoencoders", IEEE 16th international conference on data mining workshops (ICDMW), pp.241–246, 2016.
6) S. M. Lavalle, ”Rapidly-exploring random trees: A new tool for path planning”, Computer Science Dept. 98(11), 1998.