ロボティクス・メカトロニクス講演会 2016 in Yokohama

谷　口　貴　一
Kiichi TANIGUCHI
機械システム工学専攻修士課程　1 年

1. はじめに

私は 2016 年 6 月 8～11 日に神奈川県のパシフィコ横浜で行われた「ロボティクス・メカトロニクス講演会 2016 in Yokohama」に参加した．9 日のポスター講演の「進化・学習とロボティクス」セッションで「強化学習に基づく 4 脚走行ロボットのトロット歩容獲得と動作改善」について研究発表を行った．

2. 研究内容

2.1　研究背景

近年，ロボットについて盛んに研究されている．しかし，実際にロボットを制御する際には 1 つの課題がある．それは，不測の事態や未知の環境には従来の制御に少し手を加える程度では完全に対応することはできないということである．この問題を解決する手段の 1 つとして，強化学習が挙げられる．

本稿では 4 足動物をモデルとしたロボットに強化学習を適用し，トロット歩容動作の獲得・改善が可能かを調べる．

2.2　4 脚走行ロボット

本研究では，4 足動物を模した 4 脚走行ロボット（Fig. 1）をモデル化し，それをコンピュータ上に再現して学習実験を行う．ロボットの諸元は Table 1 のようになっている．

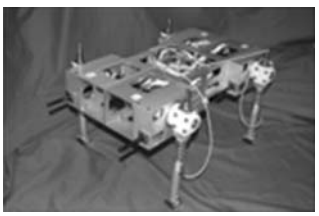

Fig. 1　Four-Legged Robot

Table 1　Detail of Robot

Length［mm］	550
Width［mm］	414
Height［mm］	100
Leg length［mm］	250
Weight［mm］	12.5

2.3　トロット歩容

動物は脚を一定の法則に基づいて動かし，それを繰り返すことで歩行や走行などの移動を行う．この動作のことを歩容という．本研究では 4 足動物が行う歩容の中でも，特にトロット歩容に着目して実験を行う．トロット歩容は，対角の脚を一対とし，それぞれの対が交互に遊脚と支持脚を入れ替えながら進む動歩行である．歩行状態から速度が増すにつれ，支持脚数が減少し，4 脚のどの脚にも体重がかかっていない滞空期間が存在する走行状態へ移行する．

2.4　強化学習

強化学習とは機械学習法の 1 つであり，環境及び状況に応じて試行錯誤を繰り返すことで目標に対して最適な行動パターンを見つけ，評価し，学習していく手法である．本研究では学習手法として Q 学習，学習方策として ε-greedy 法を使用し，学習を行う．

2.4.1　本研究における Q 学習

本研究ではロボットに「転倒」,「進行方向のずれ 35 deg 以上」,「脚振り回数 150 回以上」のいずれかが生じた場合に「失敗」として負の報酬を渡し，「目的地（5 m 地点）到達」した場合に「成功」として正の報酬を渡し，その走行中の行動を評価した．また，$\alpha = 0.4$, $\gamma = 0.9$ で学習を行った．

2.4.2　本研究における ε-greedy 法

本研究では $\varepsilon = 0.4$ で学習を行った．この方策を用いる利点は 2 つある．1 つは Q 値の高い行動のみを選択する場合と比べて多くの種類の行動を選択しうることにより探査がより広く行えることである．もう 1 つはランダムに行動を選択する場合と比べて学習中も比較的高い報酬を得やすく動作の獲得・改善に必要な行動の習得が速いことである．

2.5　学習実験

2.5.1　実験・評価概要

本研究では 4 脚走行ロボットに強化学習を適用し，学習試行させることでトロット歩容動作の獲得・改善が可能か調べた．また，本研究ではロボットに「トロット歩容のように対角の脚の対が同じ動きをし，他方の対と半周期ずれて脚をふる」という探査空間内で「脚振り周期」・「脚振り角度」の 2 つの動作パラメータについて学習を行った．実験には物理演算エンジンとして「Open Dynamics Engine : ODE」を用い，シミュレータ上で学習を行った．そして，本実験では 5 m 間を走行するという目標を設定し，3.2 節で述べた正の報酬については走行速度や消費エネルギーから算出して報酬を与えた．その際，速度に関する報酬と消費エネルギーに関する報酬について，速度報酬とエネルギー報酬のどちらも与える（(2) 式）パターンと，速度報酬のみ与える（(3) 式）パターンと，エネルギー報酬のみを与える（(4) 式）パターンの 3 パターンで比較し，報酬の与え方について検討を行った．

$$r_t = r_s + r_v + r_e \quad \text{or} \quad r_f \tag{2}$$

$$r_t = r_s + r_v \quad \text{or} \quad r_f \tag{3}$$

$$r_t = r_s + r_e \quad \text{or} \quad r_f \tag{4}$$

※r_s：到達報酬，r_v：速度報酬，r_e：エネルギー報酬，r_f：失敗報酬

2.5.2　評価結果

強化学習による動作獲得・改善実験を行った結果を図 2～4 に示す．図 2 は獲得した走行動作の一部（色の濃い脚が接地中の脚，色が薄い脚が遊脚）を，図 3 は学習中の Q 値を用いて学習せずに走行を行った（以下テスト走行と呼ぶ）際の平均速度を，図 4 はテスト走行を行った際の消費エネルギーを表している．

図 2 ではロボットがトロット歩容を行っていることがわかる．図 3 では速度とエネルギー両方の報酬を与えた場合に平均速度が 0.960 m/s（学習初期）→1.257 m/s（学習後期）と 31% 速くなり，最も学習効果が出ている．図 4 では消費エネルギーが

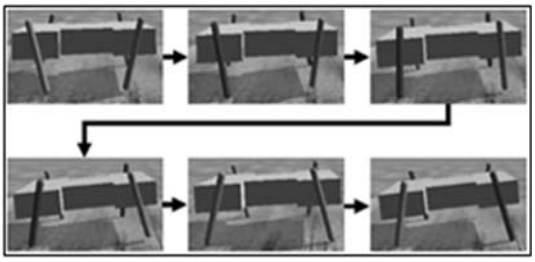

Fig. 2　Movement Robot Acquired

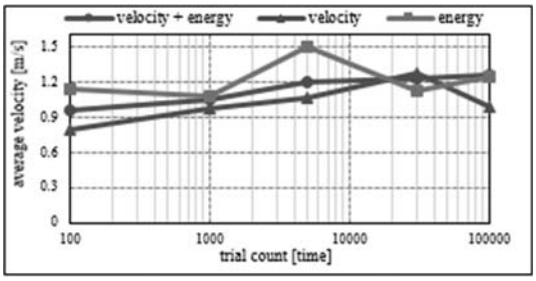

Fig. 3　Velocity（in test run）

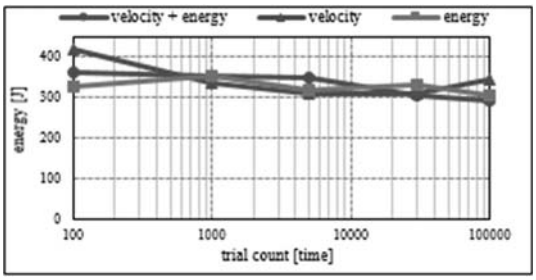

Fig. 4　Energy（in test run）

362.2 J（学習初期）→293.4 J（学習後期）と 23% 削減され，最も学習効果が出ている．これらの結果から，ロボットの走行速度と消費エネルギーは完全に同じ要因から変化が生じるわけではないため，速度とエネルギーの両方に関する報酬を与えた方が良いとわかった．

3.　まとめ

ポスター発表ではポスター以外にも動画を再生し，ロボットの動きや学習の様子をより分かりやすく伝えることができた．他の参加者や企業の研究や自分の研究に対するいろいろな意見を知ることができて，貴重な体験をすることができた．