

# 四足歩行機能を有する車両型ロボットの自律移動

柴田 泰雅<sup>†</sup>長尾 確<sup>‡</sup>名古屋大学 大学院情報学研究科<sup>†‡</sup>

## 1. はじめに

労働力不足への一つのアプローチとしてロボティクス分野においては自律ロボットなどによる労働の代替が提案されてきた。自律移動ロボットを考える中で重要となってくるのが移動効率と移動の自由度である。

一般に車両型のロボットは移動効率が良い一方で、段差などの環境に依存するため移動が制限されてしまう。また、移動の自由度が高いロボットとして四脚ロボットがあるが車両型と比べると移動効率は悪い。

そのため、本研究では2種類のロボットの欠点を補完するロボットの自律移動を目的とし、環境に応じて車両型から四足歩行型に変形するロボットの自律移動をシミュレーションを通して学習する。

## 2. 階層型強化学習と変形

### 2.1 四足歩行型と車両型の分離

自律移動ロボットを学習させるにあたって関節を動かし歩容を生成する四足歩行ロボットと車輪を回転させ動作を生成する車両型ロボットでは全く異なる機構であるため、これらの動作の学習を単一の強化学習モデルを用いて行う場合、学習アルゴリズムが非常に複雑になる。

したがって、車両型と四足歩行型のモードそれぞれで学習モデルを用意しハードウェアの変形に伴いモデルを変更することで一つの複雑なタスクを四足歩行型と車両型の二つに分割したタスクの切り替えを行う。

### 2.2 階層型強化学習

本研究においては、段差乗り越えや障害物回避などのための環境認識を動作生成と共に必要がある。そのため、それぞれのロボット学習モデルで環境認識と動作生成を含んだタスクを学習する必要がある。

Deepali ら[1]は四足歩行ロボットの障害物回避タスクに対し階層型強化学習を用いて環境認識と動作生成のタスクに分けて学習を行った。二つのレイヤに分割し、高レイヤでは環境からの入力に基づき低レイヤへの入力となるステップまでの遅延時間を出し可変周期で実行される。低レイヤではロボットの動作に関する情報を入力として動作を生成し一定周期で動作する。

このような構成にすることで環境認識と動作生成を分割して捉えることが可能となり、画像などの高次元の状態空間からモータの回転などの低次元の行動を出力する方策を学習できる。

階層型強化学習は上述のように高レイヤから低レイヤに対してメタ的な情報を出力する、そして階層型強化学習の原則としてこの出力に応じて低レイヤの挙動は変化しなければならない。

### 2.3 提案手法

階層型強化学習の種類の一つとしてオプションフレームワークというものがある。これは図1のように複数の異なる挙動を学習させた低レイヤ（オプション）を高レイヤによって切り替えることで、高レイヤは環境に応じて挙動の切り替えを学習し、低レイヤでは特定の限定された挙動のみを学習することで複雑なタスクを学習可能にするアルゴリズムである。

本研究では環境に応じて変形を行うため、このアルゴリズムを用いて四足歩行型と車両型の挙動を学習させた低レイヤを、高レイヤによって環境に応じ切り替える手法を提案する。

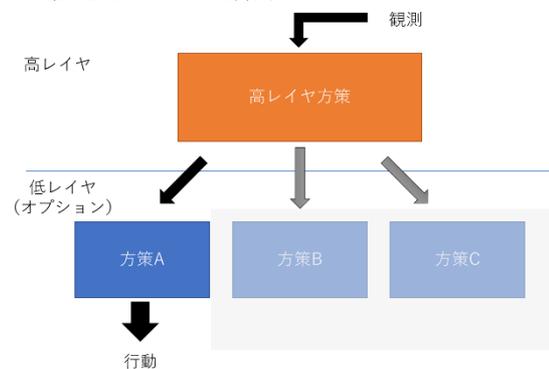


図1. オプションフレームワーク

## 3. 四足歩行モードと車両モードの共通化

本研究では物理シミュレータ Pybullet を用いてシミュレーション環境を構築し、障害物回避の階層型強化学習を行った。

### 3.1 階層型方策

各レイヤの階層型方策について述べる。

#### 3.1.1 高レイヤの構成

階層型強化学習は、それぞれのレイヤに入力する観測データの種類やレイヤ間の受け渡す情報の次元などの設計に基づき、学習した階層型方策のそれぞれのレイヤの機能は大きく変わってくる。

したがって四足歩行と車両のどちらのモードにお

いても高レイヤを同じネットワーク構成および同じ入出力にしてナビゲーションの機能を持たせる設計にすることで、高レイヤから低レイヤへの出力は同じ表現空間上のベクトルを得られると考えた。

そのため障害物や段差の認識のため、深度カメラの16x16x1の深度画像を3層の畳み込み層のCNNに入力し5次元の特徴ベクトルにした後、目標方向を決定するためのロボットとゴールの相対角度と結合して全結合層に入力し、低レイヤへの-1~1の範囲に制限した2次元の入力（以降、潜在ベクトルと呼ぶ）と次の実行までのステップ数を出力とする。

### 3.1.2 四足歩行モード低レイヤの構成

四足歩行の学習では周期的な歩容を学習する必要があり、ルールベースでは周期的な動作に向いている一方で外的要因に弱い、ニューラルネットでは周期的な動作の生成が難しい一方で転倒などの外的要因に対して柔軟に対処が可能である。

そのため四足歩行の低レイヤにおいて Atil ら[2]のルールベースの Trajectory Generator (TG)とニューラルネットを組み合わせた Policy Modulating Trajectory Generator (PMTG)を用いて学習を行う。

PMTG では TG は事前に定義された周期的な歩容生成器であり、モータの角度を出力する。そしてニューラルネットでは IMU などのセンサ情報と高レイヤからの入力および TG の現在の周期から TG のパラメータとモータの角度を出力する。そして、TG とニューラルネットの二つのモータ角度を合成することで低レイヤの出力とした。

### 3.1.3 車両型モード低レイヤの構成

車両型モードの入力として、車両の角速度と角度と高レイヤからの出力を利用し、2つのモータに対して速度指令を出力するニューラルネットとして低レイヤを構成した。

## 3.2 学習環境

学習環境は、四方13mの壁に囲まれており、ランダム性を持たせるため障害物がエピソード毎に位置が変化するように設置する。また、四足歩行モードの学習においてはこの環境に追加して高さ3cmと5cmの段差を環境の中心地点を囲むように2点設置して必ず段差を超える状況が発生する環境とした。

エージェントは環境の中心からスタートしその時の向きはランダムに0~360度に決められる。ゴールは四方の隅の内一つが選択される。

## 3.3 考察

シミュレーション実験の結果から両方のモードの階層型方策は高レイヤが潜在ベクトルを介して方向と速度を指定するコントローラとして機能しており、低レイヤはその指令通りに動作することを旨とするコントローラとして機能していることが分かった。

## 4. 変形を含めたロボットの学習

前節の分析結果を基に変形を考慮した方策のモデルを設計しシミュレーションで検証を行った。

### 4.1 提案する階層型方策モデル

どちらのモードにおいても高レイヤはナビゲーションの機能を学習し、その潜在ベクトルは同じ表現空間上のベクトルであることが前節の分析により判明した。したがって提案する階層型方策モデルはオプションフレームワークをベースとして高レイヤは前節の高レイヤの構成を用いる。オプションとして両モードの低レイヤを用いて図2に示す構成の階層型方策とした。

また、高レイヤの状態として現在のエージェントのモードを追加し、出力として変形信号を追加する。オプションはその信号に基づきハードウェアの変形と共にモデルの切り替えを行う。

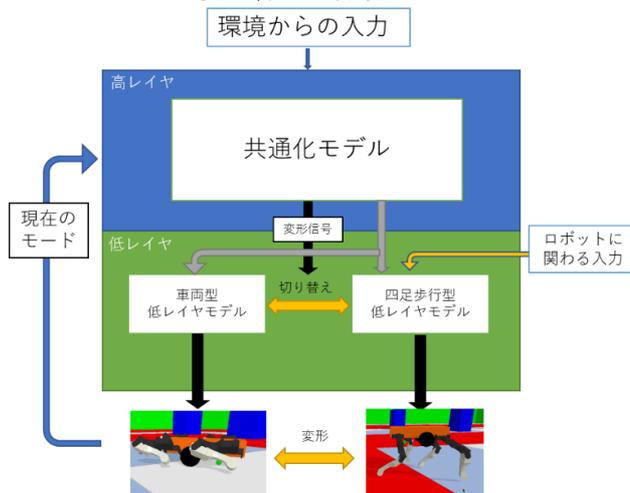


図2. 提案する階層型方策モデル

### 4.2 高レイヤの事前学習

前節の学習済みの障害物回避モデルを用いてシミュレータ内で動作させ、高レイヤの入出力をモード毎に50万サンプルずつ収集し合計で100万サンプルとした。そして、モードと変形信号のラベリングを行いそのデータをエキスパートデータとして模倣学習を行い共通化モデルとした。

### 4.3 低レイヤとの統合

模倣学習で得られた高レイヤと学習済みの低レイヤを統合する。統合後、シミュレータ内で強化学習を行い最適化した。

## 5. おわりに

本研究では、四足歩行機能を持つ車両型ロボットにおける自律移動のための階層型方策モデルを提案した。シミュレーション環境においては有効性が確認されたため、今後、実環境において実験を行い、その有効性を検証する。

### 参考文献

[1] J. Deepali, A. Iscen, and K. Caluwaerts. "From pixels to legs: Hierarchical learning of quadruped locomotion." arXiv:2011.11722, 2020.  
 [2] Atil Iscen, Ken Caluwaerts, Jie Tan, Tingnan Zhang, Erwin Coumans, Vikas Sindhwani, and Vincent Vanhoucke. Policies Modulating Trajectory Generators. CoRR, Vol. abs/1910.02812, 2019.