

# AIによる制御設計の原理研究

芝原 杏奈

## 1 はじめに

昨今、ヒューマノイドロボットの制御や自動車の自動運転などに見られるように、世間ではAI(Artificial Intelligence)<sup>注1)</sup>を用いた制御が急速に普及している。AIブームのスタート直後である約10年前、AIによる制御に関するPoC(Proof of Concept)の結果が様々な会社や団体から発表されたが、当時はシンプルで単機能のものがほとんどであった。例えば、ロボットアームによる多様な物体の把持などがあり、それらの多くはそれ以前の技術では難易度が高いが、その機能単体では製品やサービスとして成立し難いタスクであった。

その後、AIによる制御は高度化、複雑化が進み、現在では既に製品化やサービス化も始まっている。以前は実現困難であった様々な処理が現実的になりつつあり、例えば、産業用ロボットに対して少ない手間で作業内容を学習させて現場に投入することや、自動車の自動運転において、人間による制御設計の

みでは網羅が困難な多様なエッジケースに対する処理能力を高めることなどが可能になってきている。このように、近年目覚ましい速度で発展しているAIによる制御設計の技術をカヤバも獲得・応用し、制御開発の高度化や効率化につなげる必要があると考える。

そこで本研究ではAIによる制御設計の技術獲得のため、当社製品のひとつである四輪車向けセミアクティブサスペンション(以下セミアク)を対象とした原理研究を実施した。本報はその取り組み内容について紹介する。

注1) ここではAI=機械学習として説明する。

## 2 本研究の実施内容

本研究の実施内容の概要を図1に示す。AIとの比較対象となる手法にはスカイフック制御を用いた。よって、これ以降は既存のセミアク制御をスカイフック制御として説明する。セミアクで発生させる

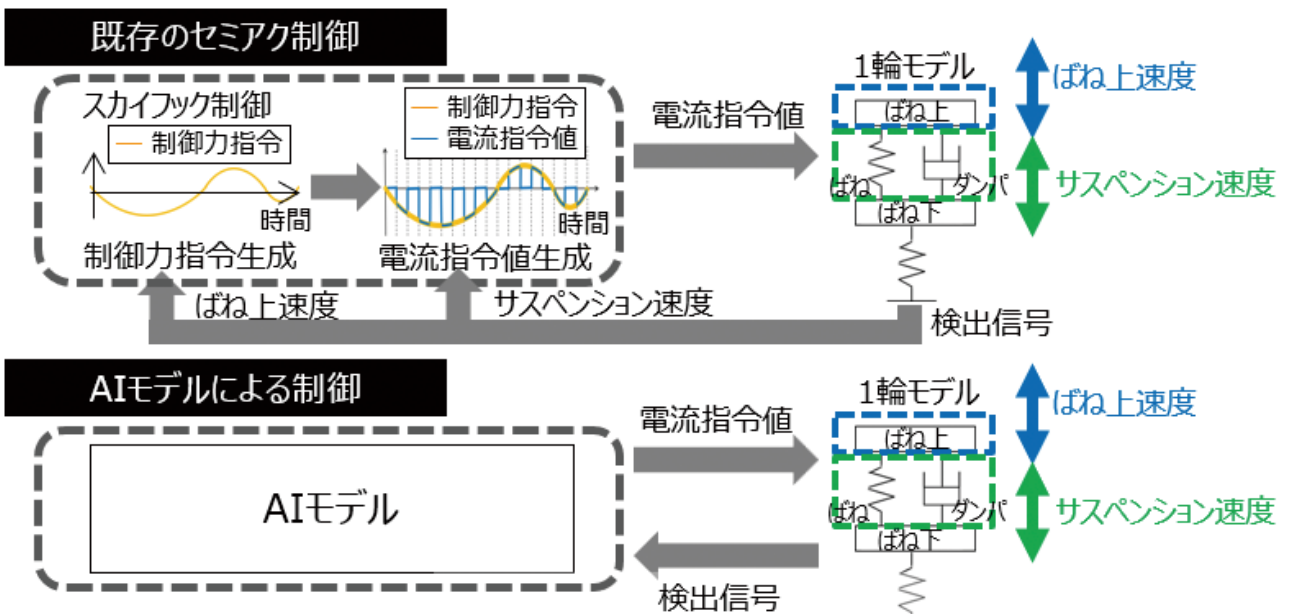


図1 本研究の実施内容

ことができる力はダンパの伸縮に対する抵抗力（減衰力）のみである。そのため、ばね上を制振する方向に力を出すことができないタイミングが存在し、そのようなタイミングにおいては最小の減衰力に制御することが望ましい。このように、セミアク制御は非線形な制御であるため、制御開発の難易度が高く、AIを導入するメリットを享受し易いと考えられる。更に、難易度の高い制御設計をAIで実現できるということは、それ以外の制御設計もAIで実現できる可能性が高いと見込まれる。

本研究ではAIによる制御設計の実現可能性を、スカイフック制御と同程度かそれ以上の性能をAIによる制御設計（AIモデル）で得られるかという観点で評価した。今回の取り組み内容は全てシミュレーション上で実施し、制御を行う対象は1輪モデル、制御が1輪モデルに与える制御値はセミアクのダンパに与える電流指令値とした。電流指令値とは、セミアクのダンパの減衰力を決定する電流の値である。

スカイフック制御の検出信号にはサスペンション速度とばね上速度を用いる。ばね上速度を用いて制御力指令を生成し、制御力指令とサスペンション速度から電流指令値を生成する。その電流指令値によって制御対象である1輪モデルを制御する。

これに対し、AIモデルによる制御は、スカイフック制御の点線で囲った範囲をAIモデルが担うものとなる。つまり、AIモデルは制御対象（1輪モデル）から検出信号を受け取り、電流指令値を返す。このAIモデルによる制御の性能を既存のセミアク制御と比較することによって評価する。

わる手法である。

このような手法に該当する代表的なAI手法群としては教師あり学習が挙げられる。教師あり学習においては、学習用データセットによって学習を行うことにより、AIモデルへの入力データに対応するべき値を出力することができる。教師あり学習に必要な教師データを用意するには多くの場合、人手を介する必要がある。そのようなデータセットによって学習したAIモデルは、基本的には人手で設計した制御と類似した制御設計となり高性能化は困難である。更に、必要な作業プロセスの観点を考えると、人手での制御設計に加えてデータセット作成とAIモデルの学習の分だけ実施すべきプロセスが増えてしまう。このように、人手での制御設計を学習に用いると、AIを用いるメリットが見出しづらい可能性がある。そこで、本研究ではAIモデルの学習に人手による制御設計が関わらないAI手法として、強化学習に分類される手法を用いることとした。

強化学習の基本的な考え方は、試行錯誤を繰り返して行うことにより経験を蓄積し、その経験から最適な行動の方策（今回で言えば制御に相当する）を獲得するというものである。強化学習の大まかな仕組みを図2に示す。

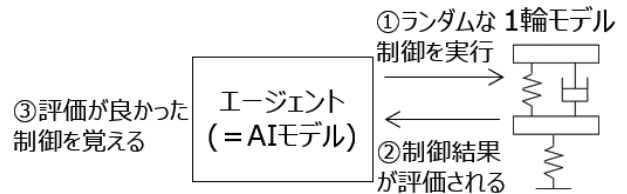


図2 強化学習の仕組み

### 3 AI手法

#### 3.1 強化学習

AIには様々な手法が存在し、手法によって前提となる問題設定や用途が異なる。この問題設定に対応する手法群を学習のパラダイムと考えると、AI手法は3種類の学習パラダイムに大別することができる。3種類はそれぞれ、教師あり学習、教師なし学習、強化学習と呼ばれる。各学習パラダイムに属する手法群には制御を学習することが可能な手法も複数含まれており、2章にて示したAIモデルを実現するためのAI手法には、学習によって制御を獲得することが可能な手法を用いる必要がある。しかし、制御の獲得が可能であればどのようなAI手法であっても良いわけではなく、実用化を見据えると使用するAI手法には向き不向きが存在すると考えられる。不向きな手法とは例えば、AIモデルの学習に用いる何らかの要素に人手による制御設計が関

今回対象とするセミアク制御という問題設定に対応する強化学習用語は次のようになる：

- (1)制御 = 方策
- (2)制御結果の出力（今回の場合は電流指令値） = 行動
- (3)制御対象（プラント、今回の場合は1輪モデル） = 環境
- (4)環境を観測して得られる検出信号（今回の場合は1輪モデルのシミュレーション結果から得られるばね上速度とサスペンション速度） = 状態

ここからは強化学習の仕組みについて説明する。強化学習において行動を実行する主体となるものをエージェントという。学習の流れとしては、はじめにエージェントがランダムな方策による行動を実行し、次にその行動を環境に適用した結果として得られる状態を評価する。この評価結果を報酬という。エージェントは可能な限り報酬が多く得られる方策

を覚える。覚えた良い方策とランダムな方策を織り交ぜながら学習を続けることで、方策が洗練されていく。良い方策とランダムな方策はそれぞれ探索と利用と呼ばれる。探索とは試行錯誤を実現するための動作であり、一方の利用はその時のエージェントにできる最も良い制御を行うための動作である。この探索と利用は同時に実行することはできずトレードオフの関係にあるため、これら2つをバランス良く実行しながらより良い方策を獲得する。

この強化学習の仕組みを実装するために必要な学習プログラムの概念図を図3に示す。学習プログラムとして設計すべき事項は3つ存在する。1つ目はエージェントの行動結果を評価する部分であり、報酬関数と呼ばれる。この報酬関数の設計によってどのような制御が学習されるかが決まる。2つ目はエージェントへの入力であり、特徴量抽出（入力信号の種類の選定と、エージェントが学習しやすいように信号を加工する）を設計する必要がある。3つ目はエージェントである。今回は強化学習手法の中でも方策の表現にニューラルネットワークを使用する手法である深層強化学習を利用した。ニューラルネットワークの構成によって学習の進み易さや性能の限界が決まるため、これを決定する必要がある。

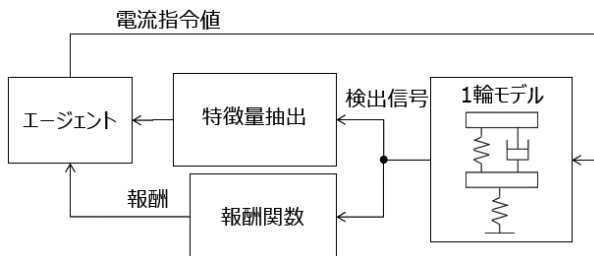


図3 強化学習の学習プログラムの概念図

### 3.2 開発環境

本研究を実施するに当たり、実装が必要となる要素は1輪モデルのシミュレーションと強化学習である。これらは全て制御モデルの構築に用いられることの多いソフトウェアであるMATLAB<sup>®</sup>とSimulink<sup>®</sup>によって実装した<sup>注2)</sup>。

注2) MATLAB<sup>®</sup>, Simulink<sup>®</sup>はMathWorks社の登録商標である。

### 3.3 使用した学習アルゴリズム

学習アルゴリズムには深層強化学習の一種であるPPO<sup>1)</sup> (Proximal Policy Optimization) を使用した。PPOとは2017年にOpenAIが発表した、学習の安定性と効率性を両立させるために設計されたアルゴリズムである。

PPOはActor-Critic法という手法に基づいている。

これは行動を実行するためのモデル（ここではニューラルネットワーク）であるActorと、行動の価値（ある状態においてその時の方策で行動した場合の累積報酬の期待値）を評価しActorに提供するモデルであるCriticの2つのモデルを用いる手法である。そのため、アルゴリズムにPPOを用いる場合にはエージェントは2つのニューラルネットワークを持つ。

今回、行動空間（取り得る行動の集合を定義するための数学的構造としての空間）は連続とし、Actorの方策はガウス分布に従うとした。

### 3.4 学習プログラムの設計

3.1節にて説明した3つに関して様々な設計を試した結果、次のようになった。

#### 3.4.1 報酬関数

報酬関数は直近の固定区間長のばね上加速度値群の最大値が小さくなれば高い報酬となるような関数とした。

#### 3.4.2 特徴量抽出

エージェントへの入力となる特徴量にはばね上速度とサスペンション速度を使用する。これら2つの値のスケール調整を行い、更にそれらの過去の値を追加してエージェントに入力することとした。

#### 3.4.3 エージェント

エージェントのネットワーク構成はActorとCriticどちらも同じ構成とし、中間層3層（要素数は128, 64, 32）とした。

## 4 学習の実施と性能評価

### 4.1 学習の進捗結果

学習時の収益（1エピソード<sup>注3)</sup>の報酬の合計）の推移は図4のようになった。このグラフより、収益は安定的に向上した後、収束に近い状態となったことが分かる。よって、学習は安定的に進捗したと言える。

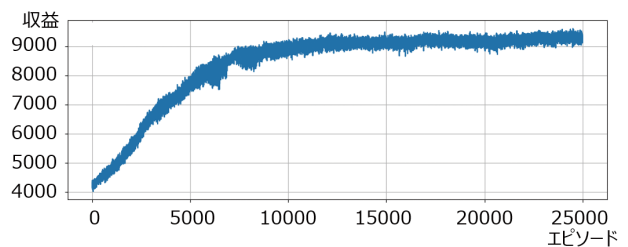


図4 学習時の収益の推移

注3) エピソード：強化学習において学習の反復数を表す単位。

### 4.2 評価方法

4.1節にて示した進捗を経た際の学習によって学習済みAIモデル（3.1節で説明したエージェントに相当）が得られた。本研究においては、この学習済みAIモデルによって推論<sup>注4)</sup>した結果、得られるデータ（ばね上速度値群）を用いてAIモデルの性能を評価する。本研究におけるAIモデルの評価方法を図5に示す。本研究では、特定の周波数帯域における振動の大きさの総和を表す値であるPOA (Partial Over All) を用いて性能評価を実施した。評価の実施手順としては、まずシミュレーション上で評価用路面を走行しつつ、学習済みのAIモデルで推論する。その結果として得られるばね上速度値群から周波数ごとの振動の大きさであるPSDを求める。また、AIモデルと性能比較するために、スカイフック制御と電流指令値固定（制御を行わない）の条件で同様にシミュレーションを実施し、その結果に対してもPSDを求める。このようにして得られたPSDから低周波領域、中周波領域それぞれのPOAを求める。得られたPOAを低周波と中周波の2軸でプロットして性能比較する。図5の右側に記載しているPOAプロット図において、赤色で示した領域が電流指令値固定のPOA、青色で示した領域がスカイフック制御のPOAである。電流指令値固定はソフト減衰力、ハード減衰力、ソフト減衰力とハード減衰力の中間の3点をプロットしており、スカイフック制御はスカイフック制御ゲインを変化させて12点（ゲイン-3000, -5000, -10000, -20000, ……,-100000）プロットしている。POAプロット図の左下になるほど高性能となるため、AIモデルは評価結果のPOAが可能な限り左下に位置するように学習プログラムを設計している。

注4) 推論：ここでは、方策更新・探索を行わずにセミアクを制御すること。

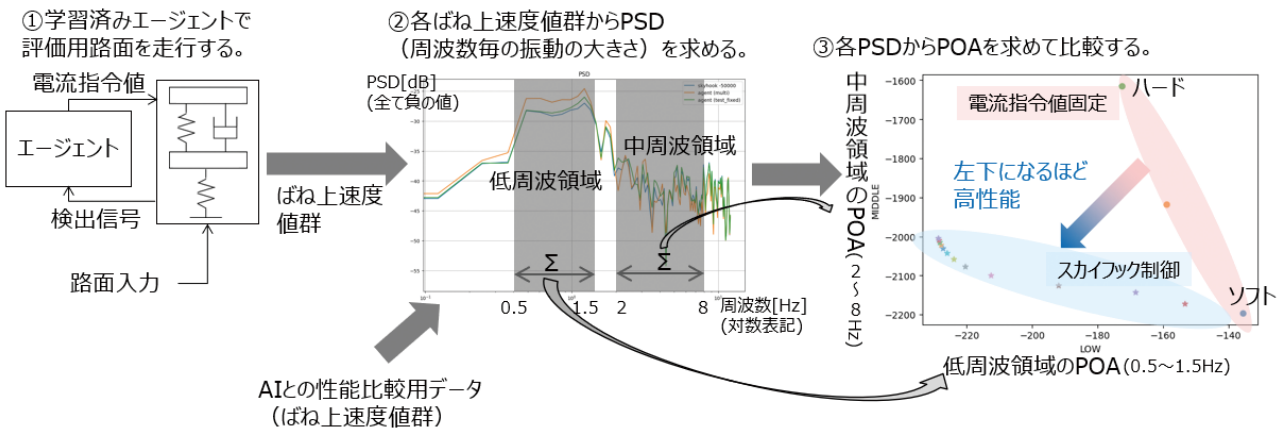


図5 AIモデルの評価方法

### 4.3 評価結果

AIモデル、スカイフック制御、電流指令値固定のばね上速度のPOAを図6、AIモデルと最もPOAが近かったスカイフック制御（ゲイン-30000）とAIモデルの電流指令値の推移を1秒分抜粋したものを図7に示す。AIモデルのPOAは比較対象であるスカイフック制御のPOA値群の内、ゲイン-30000の場合のPOAとほぼ同等となっていることがわかる。また、AIモデルが出力した電流指令値はスカイフック制御に近い推移を取るといった結果となった。この結果より、AIモデルは制御性能、電流指令値の推移ともにスカイフック制御に類似した制御挙動を示したと言える。すなわち、AIによる制御設計は既存のセミアク制御と比較して十分な性能の制御を実現可能な見通しが立ったと考える。

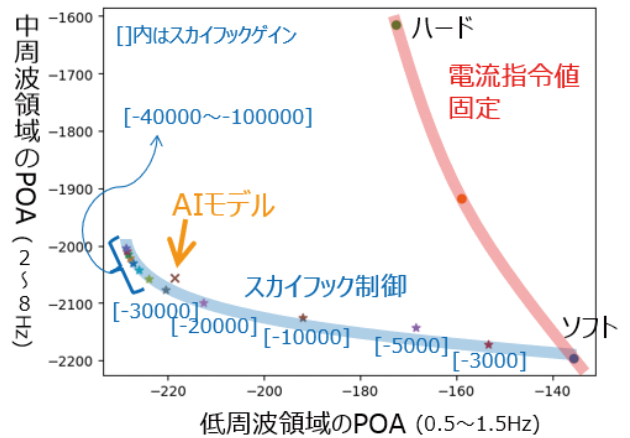


図6 AIモデル、スカイフック制御、電流指令値固定のばね上速度のPOA

なお、今回の取り組みにおいては1種類の路面のみを使用しているため過学習となっている懸念があるが、その点に関しては今後の取り組みにて制御の

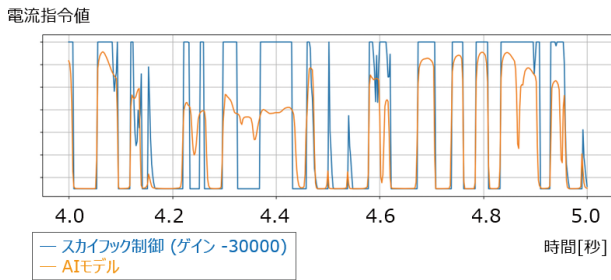


図7 スカイフック制御とAIモデルの電流指令値推移

適用対象に応じた多様なシーンへの対応を行う予定である。

## 5 おわりに

セミアク制御設計を対象として、AIによる制御設計の原理研究を行った。その結果、比較対象であるスカイフック制御とほぼ同等の性能かつ、スカイフック制御と類似した電流指令値を出力するAIモデルが得られた。この結果より、AIによる制御設計の基本的な技術獲得ができたと考えられる。

今回の取り組みは限定的な範囲のものであったが、今後はより実用的な取り組みに移行し、実際の制御設計への本技術の活用を目指す所存である。

### 参考文献

- 1) 論文投稿サイトarXivのPPOに関する文献掲載ページ  
<https://arxiv.org/abs/1707.06347>

## 著者



芝原 杏奈

2015年入社。技術本部基盤技術研究所運動制御研究室。AIによる制御の研究開発に従事。