

# 強化学習を用いた PMSM の駆動ロジック自動構築手法

Method to Automatically Generate PMSM Speed Control Model Using Reinforcement Learning

茂田 智秋 SHIGETA Tomoaki 高野 俊也 TAKANO Toshiya

機械学習の手法である強化学習 (RL : Reinforcement Learning) では、制御対象の行動を試行錯誤的に変化させ、より良い状態となるような駆動ロジックを自動で構築できる。産業分野でモータトルクや速度精度が重要視される負荷では、駆動条件ごとに制御の調整が求められるので、RL の適用によって調整時間の短縮が可能になる。

そこで東芝インフラシステムズ(株)は、制御のリアルタイム性が重要となる制御対象として永久磁石同期電動機(PMSM)をモチーフに、RLを実時間で行うための駆動ロジック自動構築手法を開発した。コンプレッサーを模擬した負荷条件で、従来の比例積分 (PI) 制御と比べ速度脈動を 1.6 ポイント抑制できることを確認した。

Reinforcement learning (RL) is a machine learning method that makes it possible to automatically acquire an enhanced control strategy by changing the behavior of the controlled object through trial and error. As it is important for motors in the industrial field to meet requirements for torque and speed accuracy, the application of RL is considered to be an effective means of shortening the time required to change control models that depend on individual driving conditions.

Toshiba Infrastructure Systems & Solutions Corporation has developed a method to automatically generate a speed control model using RL in real time, using permanent magnet synchronous motors (PMSMs) as a case study. From the results of experiments using an actual PMSM together with a motor that can simulate the load conditions of a compressor, we have confirmed that this RL-based speed controller makes it possible to suppress speed ripples by 1.6% compared with the conventional method using proportional-integral (PI) control.

## 1. まえがき

近年、自動運転技術やロボット制御技術など、モデルが複雑で高度な制御が要求される分野におけるブレークスルー技術として、人工知能技術の一つであるRLが目ざされている。RLは、図1(a)に示すように教師あり学習及び教師なし学習と並ぶ、機械学習の手法の一つとして位置付けられている。RLは、図1(b)のように“エージェント”と“環境”から構成され、エージェントは環境の状態を観測し、その結果を基に行動を決定する。環境は、与えられた行動に従って新たな状態に移り、その行動に対する報酬を与える。このようなやり取りを繰り返すことで、エージェントは様々な状態に対応した適切な行動を学習していく。更に、RLは直接正解を与えて学習する教師あり学習とは異なり、報酬を指標として行動を学習する。このように、環境に関する完全な理解を必要としないことから、複雑な制御系への応用が期待できる<sup>(1)</sup>。

一方、東芝インフラシステムズ(株)は、カーボンニュートラルの推進に向け、高効率なPMSMの市場展開を進めている<sup>(2)</sup>。PMSMの性能を十分発揮させるためには、負荷や周囲環境の変化による機器の状態を考慮した制御調整が必

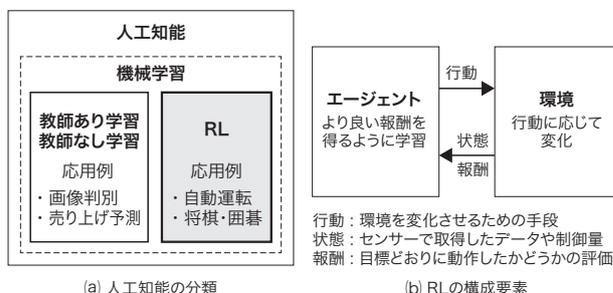


図1. RLとその構成要素

RLは直接正解を与えて学習する教師あり学習とは異なり、報酬を指標として行動を学習する。複雑な制御系への応用が、期待されている。

### RL and its components

要であるが、そのような調整は多大な労力を要することがあり、改善が望まれている。

そこで当社は、このような課題を解決する方法としてRLに着目し、駆動ロジックを自動構築するアルゴリズムを開発した。ここでは、考えられる多様な応用先の一つとして、トルク脈動を発生するコンプレッサー負荷に対してPMSMの速度制御を行うケースをモチーフとし、RLをリアルタイムに行うために考案したRLシーケンスと報酬計算方法について

述べる。また、実機試験評価を行って、この計算方法の有用性を確認したので、併せて述べる。

## 2. PMSMの速度制御の問題点

一般的なPI速度制御とRLを用いた速度制御について説明する。PMSMの電圧や電流情報を基に回転角度と速度を推定演算するセンサーレス制御を採用した、PMSMの駆動用インバーター及び制御基板を、図2に示す。制御基板は、速度制御機能とベクトル制御機能を搭載し、PMSMを速度指令値どおりに駆動するためのゲート指令を生成し、インバーターの半導体をスイッチングしてPMSMに通電する。

### 2.1 一般的な速度制御方法とその問題点

PMSMの速度（回転速度）を制御するためには、トルクを適切に制御する必要がある。PMSMにおいて、回転子に同期したdq<sup>(注1)</sup>座標系のトルク $T_{rq}$ は、式(1)で表される。

$$T_{rq} = P \cdot \{ \phi_{mag} + (L_d - L_q) \cdot i_d \} \cdot i_q \quad (1)$$

ここで、モーター極対数<sup>(注2)</sup> $P$ 、永久磁石の磁束 $\phi_{mag}$ 、 $d$ 、 $q$ 軸の電流 $i_d$ 、 $i_q$ 、及び $d$ 、 $q$ 軸のインダクタンス $L_d$ 、 $L_q$ である。

図3に示すように、速度指令と検出速度の差（速度偏差）をPI制御し、式(1)の $q$ 軸電流を調整する速度制御方法が、広く利用されている。これに加えて、あらかじめ用意したテーブルデータやオブザーバーで、負荷トルクによる速度脈動を推定し、この値を用いてトルクを補償する方法も提案されている。一方で、テーブルの個別調整やゲイン調整に時間を要するといった問題がある。

### 2.2 従来のRLを用いた速度制御方法とその問題点

RLの一手法として、行動モデル(Actor)と報酬モデル

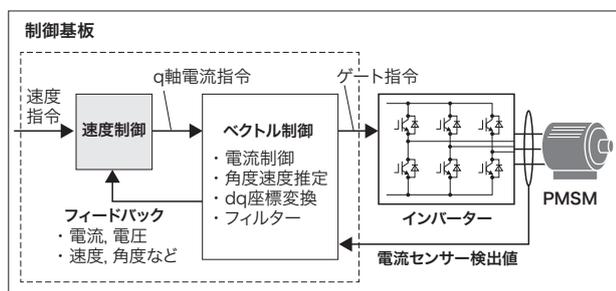


図2. PMSM駆動用インバーターと制御基板

PMSMの電圧や電流情報を基に回転角度と速度を推定演算する、センサーレス制御を採用した。

Inverter and controller for driving PMSM

(注1) モーター回転子の磁極方向を $d$ 軸、直交方向を $q$ 軸とした座標系。  
(注2) モーター磁極のN極とS極の二つを、一対とした数。

(Critic)をエージェントとしたActor-Critic法がある。この手法は、前述のテーブルデータや負荷モデルのような教師データを事前に準備する必要がなく、行動、状態、及び状態遷移に伴う期待報酬を学習し、それに応じて最適な行動を選択する。

この手法をPMSMの速度制御に適用すると、図4(a)のようになる。観測した状態（速度や電流）から報酬即値を計算し、この結果と報酬モデルの出力である報酬期待値（報酬の推定値）とが一致するように学習を進める。また、行動モデルは、状態に応じて報酬期待値が最大化する行動を出力するように学習を進める。この出力を $q$ 軸電流指令に割り当てることで、PMSMの速度を制御する構成とする。このとき、報酬・行動モデルは、それぞれ図4(b)に示すような多入力、多層のニューラルネットワーク(NN)で構成され、

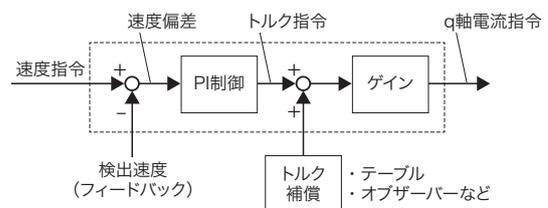
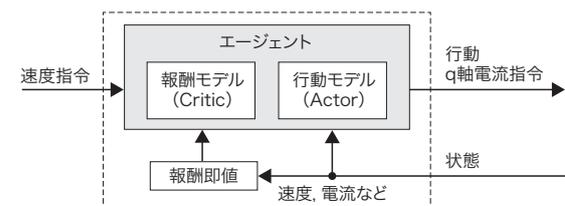


図3. PI制御によるPMSM速度制御ブロック図

一般的には、PI制御方式が広く利用されている。

Block diagram of automatic speed regulator with PI control



(a) RL手法を用いた速度制御

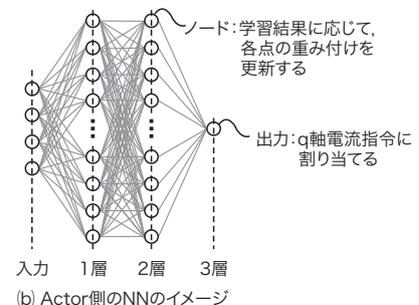


図4. RL手法を用いたPMSM速度制御ブロック図

報酬モデルと行動モデルは、それぞれ別々のNNで構成され、NNの各ノードの重み付けを計算・更新することが学習に相当する。

Block diagram of automatic speed regulator using RL method

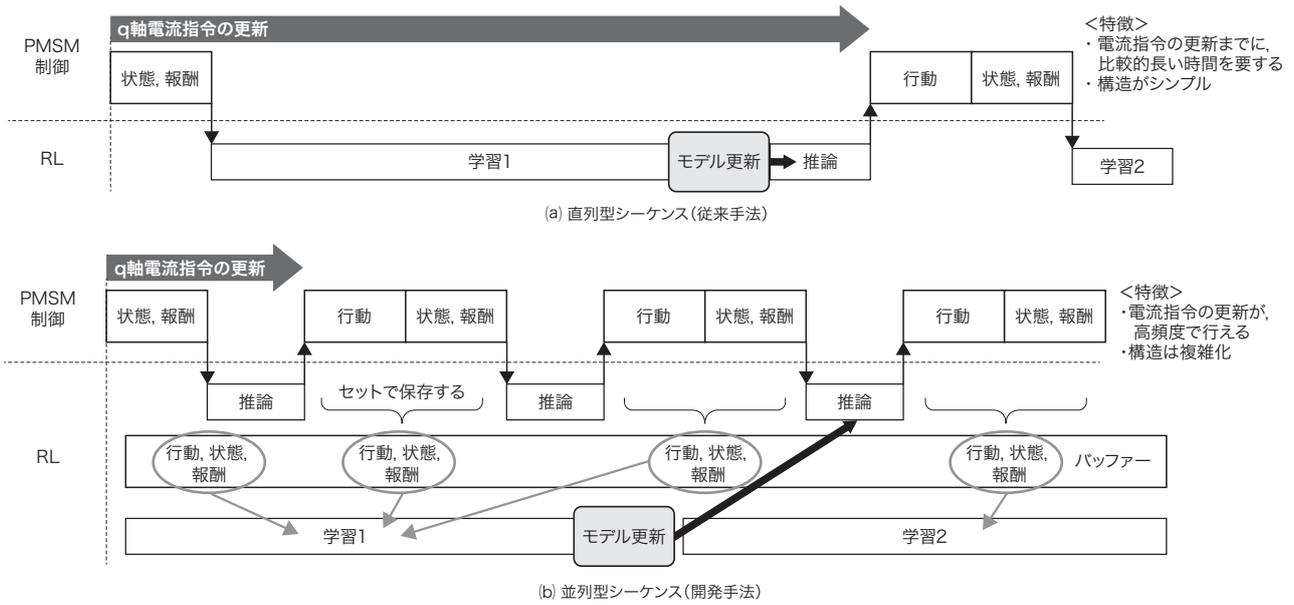


図5. RLシーケンスとq軸電流指令の更新頻度の比較

直列型では大きなq軸電流を選択した場合、次回の電流指令の更新までに比較的に長い時間を要するため、加速し続けて保護停止することがある。一方、並列型では構造は複雑化するが、電流指令の更新をより高頻度で行うことができる。

Comparison of RL sequences and update rate of q-axis current commands

NNの各ノードの重み付けを計算・更新することが学習に相当する。

一般に、NNを用いることでより複雑な対象に対応できるものの、計算に時間を要する傾向にあり、これがリアルタイム性を重視する用途にRLを適用する場合の阻害要因となっている。

### 3. PMSM速度制御へのRL適用の検討

RLを用いてPMSMの速度を制御するためには、電流指令の更新を数ms程度で行うことと、適切な報酬モデルを構築することが必要となる。先行研究<sup>(3)</sup>では理想条件でシミュレーションを行い、学習に付随する時間的な遅延や制約については考慮しなかった。

ここでは、リアルタイム性が重要となるPMSM実機へのRL適用を目的として検討した内容について述べる。

#### 3.1 RLシーケンスの並列化

PMSMを駆動するための最もシンプルな直列型RLシーケンスを、図5(a)に示す。この図ではセンサーなどで検出した状態をエージェントへ渡した後に、学習が完了するまで電流指令の更新を待機する。この場合、シーケンスは簡素かつ最新の行動モデルを利用できるものの、電流更新の頻度が低いために、PMSMの電流を精度良く制御できない。

そこで、図5(b)に示すように、比較的高速に行動を計算する推論(行動推定)と、学習モデルの更新を並行して行

う並列型RLシーケンスを考案した。推論は、現在までに学習済みの行動モデルを用い、検出した状態を用いてq軸電流指令を高速に計算する。また、状態、行動、及び報酬即値を、一つにまとめてバッファに保存する。このとき学習は、PMSM制御とは非同期のタイミングで保存データをバッファから取り出して実施する。十数msを要して学習が完了したタイミングで、行動モデルのNNの重み付けを更新し、更新後の行動モデルで次回以降の推論を行う。このような構成とすることで、電流指令の更新頻度の向上と時間を要する学習とを両立でき、リアルタイム性が重視されるPMSM制御に適用できる。

#### 3.2 速度脈動を低減するための報酬

行動・報酬モデルを自動構築するためには、制御が適切か否かを決定づける報酬即値計算が重要となる。図6(a)は、速度脈動をそのまま報酬即値とした場合について示している。この場合、速度脈動は正負に変化し、特に回転周波数が増加した場合に報酬の高/低が頻繁に入れ替わる。このため、各種データの取得タイミングや電流指令の更新タイミングを厳格に管理する必要が生じ、適用が難しくなる。そこで図6(b)のように、速度脈動の振幅(ピーク値)を検出し、報酬即値として利用する方法を考案した。振幅を用いる方法であれば報酬は正負に変化せず、その値が小さいほど高報酬であるという報酬即値計算の簡単化ができる。

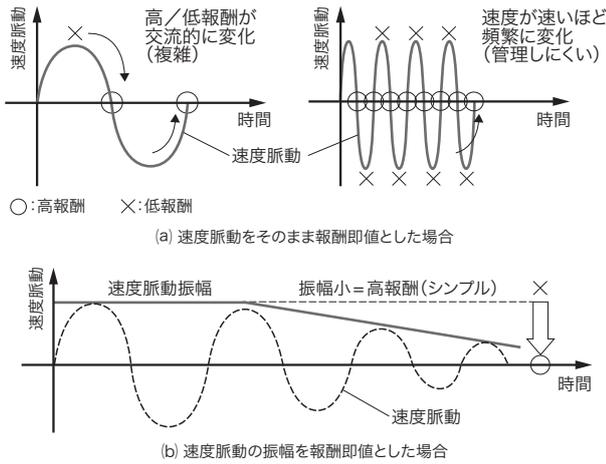


図6. 速度脈動の抑制に適した報酬

速度脈動の振幅を検出して報酬即値とし、その値が小さいほど高報酬とすることで、報酬即値計算の単純化ができる。

Rewards suitable for suppression of speed ripples

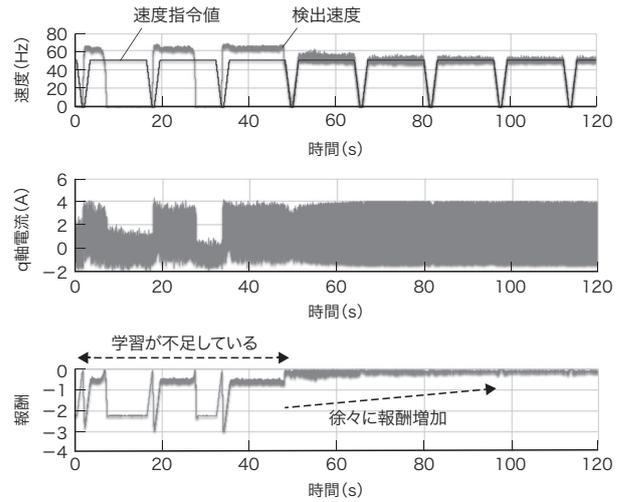


図8. 学習中の速度、q軸電流、及び報酬即値の波形データ

複雑な制御対象になるほど、モデル構築までに時間を要する。今回の例では、120 s程度でモデルが構築できた。

Waveform data of speed, q-axis current, and immediate reward under learning

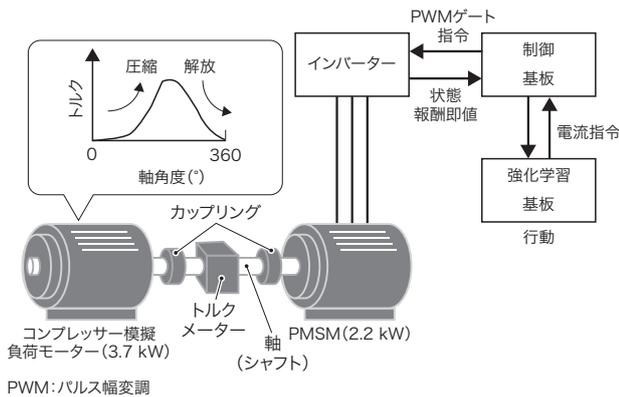


図7. PMSM駆動試験セットアップ

PMSMは2.2 kW機を、負荷モーターは3.7 kW機をそれぞれ用いた。強化学習のための追加センサーなどは、設けていない。

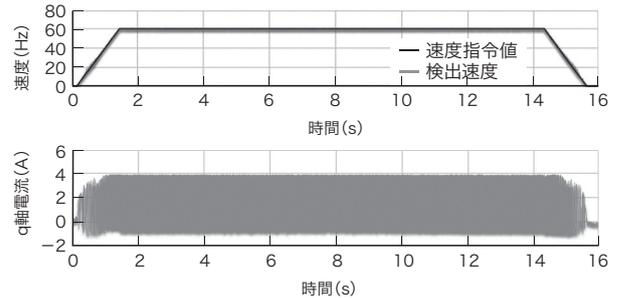
Setup of PMSM speed control experiments

#### 4. 実機試験評価

図7に示すように、コンプレッサー模擬負荷モーターとPMSMを軸直結して、評価試験を実施した。負荷モーターは、軸が機械的に1回転するとトルク脈動を1回発生するようにトルク制御を行う。このとき、検出速度の最大値 $\omega_{\max}$ と最小値 $\omega_{\min}$ を一定期間取得し、リップル含有率を式(2)で計算して、開発した手法と従来手法とを比較する。

$$\text{リップル含有率(\%)} = \frac{\omega_{\max} - \omega_{\min}}{\omega_{\text{avg}}} \times 100 \quad (2)$$

$\omega_{\text{avg}}$ : 速度平均値



(a) 学習済みモデルを用いて駆動した場合

(b) RLとPI制御のリップル含有率比較

図9. 学習完了したモデルによる試験波形と、RLとPI制御での速度脈動波形の比較

速度指令と検出速度の波形が一致しており、PMSM駆動ロジックを自動構築できている。RLを用いた制御で、リップル含有率がPI制御に比べて1.6ポイント抑制できた。

Waveform data using learned model and comparison of speed ripples using RL and PI control

従来手法のPI制御については、安定に駆動可能な条件下で、可能な限りゲインを高める調整を行った。

#### 4.1 RLの実施

図8に、実時間で120sほどの学習を実施した試験結果を示す。試験開始直後は、駆動ロジックの学習が十分できていないため、検出速度は速度指令に一致しない。しかし、学習が進むにつれて獲得できる報酬が徐々に増加し、最終的に速度指令どおりに駆動できることを確認した。

#### 4.2 完成したRLモデルの評価

学習が完成したモデルを用い、推論だけでq軸電流指令を生成してPMSMを駆動させた試験結果を、図9(a)に示す。速度指令と検出速度の波形が一致していることから、PMSM駆動ロジックを自動構築できたことが分かる。また図9(b)のように、リップル含有率をPI制御の5.5%から3.9%へと、1.6ポイント抑制できた。

### 5. あとがき

制御のリアルタイム性が必須となるPMSMの速度制御に適用可能なRLアルゴリズムを開発し、コンプレッサー負荷を模擬した試験において、速度脈動をPI制御と比較して1.6ポイント抑制した。開発したRLを応用した駆動ロジックの自動構築は、モデルに依存しないため、PMSM制御に限らず広く適用可能と考えられる。

今後は、RL技術を応用した製品展開に向けて、開発を推進する。

### 文献

- (1) 研究開発戦略センター. 研究開発の俯瞰報告書 システム・情報科学技術分野 (2021年). 国立研究開発法人科学技術振興機構 研究開発戦略センター. 2021, 520p. <<https://www.jst.go.jp/crds/pdf/2020/FR/CRDS-FY2020-FR-02.pdf>>, (参照 2021-07-07).
- (2) 東芝産業機器システム(株). “永久磁石 (IPM) モータ”. PMモータ. <<https://www.toshiba-tips.co.jp/products/motor/pm-mot/>>, (参照 2021-07-07).
- (3) 高野俊也, 齋藤亮介. “強化学習によるモータ回転速度制御のシミュレーション検証”. 2019年 電気学会 電子・情報・システム部門大会. 沖縄, 2019-09, 電気学会, 2019, p.1210-1214.



茂田 智秋 SHIGETA Tomoaki

東芝インフラシステムズ(株) インフラシステム技術開発センター  
電機応用・パワエレシステム開発部  
電気学会会員  
Toshiba Infrastructure Systems & Solutions Corp.



高野 俊也 TAKANO Toshiya

東芝インフラシステムズ(株) インフラシステム技術開発センター  
システム制御・ネットワーク開発部  
電気学会会員  
Toshiba Infrastructure Systems & Solutions Corp.