

PS-27 深層強化学習によるリニア式波力発電装置の最適制御

海洋先端技術系 * 梅田 隼、藤原 敏文、井上 俊司

1. はじめに

多くの波力発電装置 (WEC) が提案¹⁾されており、その中でも図-1に示すポイントアブソーバー型 WEC は構造的に簡便であり、制御により広範な波浪条件下で発電が可能である。本研究では、強化学習²⁾³⁾を WEC の制御への適用した。本研究では、シミュレーション上で、発電機構が永久磁石式リニア発電機であるリニア式 WEC に強化学習を適用し、学習による不規則中の発電量の変化を評価したので報告する。

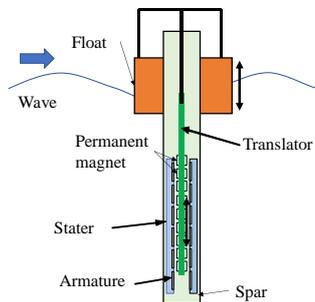


図-1 リニア式波力発電装置

2. 数値計算

波力発電装置のフロートの上下運動の方程式は、次式で表される。

$$m_f \ddot{z} = F_w + F_r + F_c + F_{ext} \quad (1)$$

ここで、 $m_f, z, F_w, F_r, F_c, F_{ext}$ はフロート質量、変位、ラディエーション力、波強制力、復原力、制御力である。このような方程式を基本とした。ラディエーション力、波強制力については境界要素法により計算した。制御力および発電量 P_e は(2)、(3)式により得られる。

$$F_c = -k_f z - k_v \dot{z} \quad (2)$$

$$P_e = F_w \dot{z} - R \left(\frac{\dot{z}}{v} \right)^2 \quad (3)$$

ここで、 k_f, k_v は制御パラメータ、 R は巻線抵抗、 K_t は推力定数である。 k_f, k_v は変化量が与えられ、(4)、(5)式により更新する。

$$k_f \leftarrow k_f + \Delta k_f \quad (4)$$

$$k_v \leftarrow k_v + \Delta k_v \quad (5)$$

対象とした WEC の主要を表-1に示す。不規則波の条件は有義波周期 1.6s、有義波高 10cm である。不規則波の初期位相は、試行 (Episode) 毎に乱数で与えて、異なる波形を作成した。 K_g, C_g の初期値は -100N/m、5Ns/m とした。計算時間は 192 s で刻み時間 Δt は 0.01 s である。制御パラメータは 0.1 s 毎に更新する。

表-1 モデル主要目

Part	Item	Unit	Value
Float	Outer diameter	m	0.3500
	Inner diameter	m	0.1140
	Draft	m	0.1500
	Mass	kg	12.90
Linear motor	Thrust	N/A	37.93
	Resistance	Ω	2.115

3. 強化学習手法

3.1 Q-learning および Deep Q-network

強化学習の一つである Q-learning では状態 s で行動 a に対して報酬 r を与えられるようなタスクに対して、実際に s で a を行って得られた r から、最終的に報酬を最大化するような連続した行動を何らかの形で学習する。このとき、行動価値関数 $Q(s, a)$ と呼ばれる関数で、現在の状態からの各行動を評価し、次取るべき行動を決定する。Q-learning では、以下の式で $Q(s, a)$ を更新する。

$$Q(s, a) \leftarrow Q(s, a) + \gamma [R + \max_{a'} Q(s', a') - Q(s, a)] \quad (6)$$

ここで、 α は学習率と呼ばれるパラメータで、 $Q(s, a)$ の更新量を制御する。 γ は割引率であり、未来の報酬をどの程度考慮するかを表す。 $\max_{a'} Q(s', a')$ は行動後の遷移状態 s' で、可能な行動 a' のうち、 $Q(s', a')$ の最大値である。矢印は代入処理を表す。右辺第二項が目標値との誤差であり、TD 誤差と呼ばれる。Deep Q-network (DQN) では Q-learning における行動価値関数 $Q(s, a)$ を多層ニューラルネットワークにより関数近似する。ただし、ニューラルネットワーク化により、パラメータが増大し、そのままでは勾配消失や過学習に陥るなど上手く学習できない問題が生じることが知られている。そこで、学習を安定的に進めるテクニック^{3,4)}が提案されている。

3.2 Huber 損失

TD 誤差を評価する損失関数 L には、二乗誤差が良く用いられるが、これは大きく値がずれた場合、その誤差が更新に大きく影響する。そこで、ロバスト回帰でよく用いられる Huber

損失を使用する。TD 誤差を(7)式とすると、Huber 損失は(8)式となる。

$$\delta = (Q + \max[\delta, \delta']) - Q(s, a) \quad (7)$$

$$\delta = \begin{cases} \frac{1}{2}\delta^2 & \text{if } |\delta| \leq \delta \\ \delta \left(|\delta| - \frac{1}{2}\delta \right) & \text{otherwise} \end{cases} \quad (8)$$

ここで、 δ は閾値であり、 $\delta=1.0$ とした。閾値の内側では2次関数だが、閾値より外側では線形増加となり、外れ値の影響が二乗誤差に比べて小さくなる。

3. 3 Dueling DQN

Q-learning では1回の更新で1つの行動に対する価値更新する。このとき、Q 関数を持つ情報は状態 s で決まる部分と行動 a で決まる部分に分離できると考えられる。そこで、図-2および(9)式で示すように状態 s だけで決まる状態行動価値 $V(s)$ と、行動で決まる部分を Advantage 関数と呼ばれる $A(s, a)$ に分けて学習し、最後の出力層で V と A を足し算して Q を求める⁴⁾。状態行動価値 V は行動によらず毎回学習できるため、早く学習が進みやすくなる。このとき、Dueling DQN で追加したニューラルネットワークの最後の層は平均値を出力から引き算する。

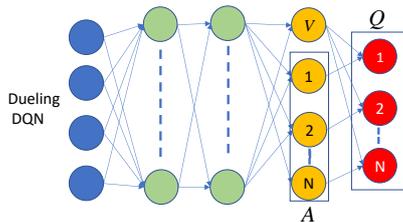


図-2 Dueling DQN のネットワーク構造

$$Q(s, a) = V(s) + \left(A(s, a) - \frac{1}{N} \sum_{a'} A(s, a') \right) \quad (9)$$

3. 4 学習モデル

変位 \hat{x} 、速度 \hat{v} 、 \hat{a}_p 、 \hat{a}_s 、波振幅 ζ (5変数)をサンプリング0.1 s、100個のデータを1フレームとして、入力データは過去5フレーム分で $5 \times 100 \times 5$ 次元のベクトルである。3層の畳み込みニューラルネットワーク、2層の全結合型ニューラルネットワーク、Dueling DQNによる結合層の後、 $Q(s, a)$ を出力し、行動を選択する。行動 a には $\Delta \hat{a}_p = +1, 0, -1, \Delta \hat{a}_s = +5, 0, -5$ の $3 \times 3 = 9$ 通りセットされ、 $Q(s, a)$ が最も高い行動を選択する。

4. 計算結果

DQN と Dueling DQN および損失関数を二乗誤差、Huber 損失を組み合わせて計算を行った。Episode 経過毎の総発電量の変化を図-3、3000Episode以降の平均値を表-2に示す。

いずれも安定して学習しているが、“Huber 損失+Dueling DQN”では、3000Episodeあたりで200Wに達しており、その他の手法よりも学習が早い。また、“Huber 損失+DQN”と“二乗誤差+Dueling DQN”の学習スピードはほぼ同じである。3000Episode以降の平均値では、“二乗誤差+DQN”の組み合わせは、3000Episodeでは学習途中であるため最も発電量は低くなった。Dueling DQN と DQN では損失関数に関わらず、Dueling DQNの方が発電量は高い。Dueling DQN 使用時の Huber 損失の効果は発電量に対してわずかである。

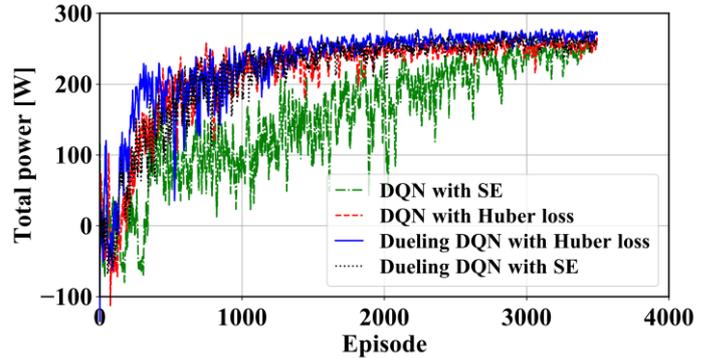


図-3 学習アルゴリズムによる発電量の比較

表-2 学習後の発電量の比較

	Square error (SE)	Huber loss
DQN	226 [W]	244 [W]
Dueling DQN	266 [W]	267 [W]

5. まとめ

リニア発電機を搭載したポイントアブソーバー型波力発電装置を対象に、強化学習の一つであるDQNを適用し、不規則波中の発電量を評価した。学習が進むことで発電量が改善することが確認できた。また、学習アルゴリズムをDueling DQNにすることで学習スピードおよび発電性能が向上する。

参考文献

- 1) 前田久明, 山下誠也: 海洋エネルギー利用<特集>1.2 波浪エネルギー一次変換装置, 日本造船学会誌, 第637号, 1982, pp. 306-326
- 2) 三上貞芳, 皆川誠章ほか: 強化学習, 森北出版, 2000
- 3) Mnih, Volodymyr, et al.: Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602 2013
- 4) Wang, Ziyu, Schaul, Tom, Hessel, Matteo, van Hasselt, Hado, Lanctot, Marc, de Freitas, Nando: Dueling Network Architectures for Deep Reinforcement Learning, arXiv preprint arXiv:1511.06581, 2015
- 5) 梅田隼, 藤原敏文, 井上俊司: 強化学習を用いた波力発電装置の最適制御, 日本船舶海洋工学学会講演会論文集, Vol. 26, pp. 129-132, 2018