強化学習を利用した波力発電装置の制御

梅田 隼*,藤原敏文*

Optimal Control of Wave Energy Converters Using Reinforcement Learning

by

UMEDA Jun and FUJIWARA Toshifumi

Abstract

This paper describes the control method we developed to maximize the output energy of a point-absorber type wave energy converter (WEC) using reinforcement learning. Conventional control methods necessitate the dynamic model of the WEC. Modeling errors of the dynamic model, however, can decrease output energy and result in inaccurate control. The proposed method does not require a dynamic model of the WEC because the method uses reinforcement learning to learn optimal control on the basis of prior experiences. We performed numerical simulations in irregular waves to compare the proposed method with the conventional method. The average output power in the proposed method was comparable to that of the conventional method. Moreover, as an advantage, the proposed method decreased power fluctuation rather than the conventional one. We verified that the proposed method was more effective than the conventional control method.

* 海洋先端技術系
原稿受付 令和 3年 1月 27 日
審 査 日 令和 3年 3月 9日

目 次

目	次	8
1.	まえがき・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	9
2.	強化学習	9
3.	運動シミュレーション ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	0
	3.1 運動シミュレーションモデル ・・・・・・ 2	0
	3.2 計算条件	2
4.	制御方法	2
5.	計算結果	5
6.	まとめ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	7
謝	辞	7
参	考文献······ 2	7

記号

a:エージェントの行動[-] r:エージェントが行動によって得た報酬 [-] s:エージェントの状態 [-] s':エージェントの行動後の状態 [-] α:学習率 [-] γ:割引率 [-] Q(s,a): 行動価値関数 [-] *m_f*:フロート質量 [kg] *m_a*:付加質量 [kg] m_∞:周波数無限大での付加質量[kg] z:フロート上下変位[m] c33:復原力係数 [N/m] fpro:リニア発電機の制御力 [N] N(ω):造波減衰力係数 [Ns/m] *E*₁(ω): 波強制力係数 [N/m] *S*(*ω*): 波スペクトル [m²s] *ω*:角周波数 [rad] P:リニア発電機の吸収パワー [W] Cg:制御パラメータ [Ns/m] Kg:制御パラメータ [N/m] P_c :制御により生じた内部損失(銅損) [W] $P_n: リニア発電機から取り出した電力 [W]$ K_t: 推力係数 [N/A] R: 巻線抵抗 [Ω] T_s :有義波周期 [s] H_s:有義波波高 [m] *ζ*:波変位 [m] X:機械リアクタンス [kg/s]

1. まえがき

波力発電装置(Wave energy converter: WEC)の1つである Fig.1に示すポイントアブソーバー型は,波により上下 に運動するフロートと発電機構が内蔵されるスパーの2要素で構成される.ポイントアブソーバー型で採用され る発電機構に様々な方式が検討されているが,海技研では発電機構に永久磁石式のリニア発電機 (PMLG)を持つ リニア式 WEC の開発を行ってきた¹⁾. リニア式 WEC のメリットは,リニア発電機の制御応答が良く,ギアなど の変換機なしに発電機を駆動できるため,機構が簡便となることである. Fig.1 の WEC の場合,スパーに PMLG の固定子を固定し,治具を介してフロートに可動子を接続する.フロートとスパーの相対運動により可動子は上 下する.可動子が固定子の永久磁石の周りを移動すると,磁界変動により誘導起電力が生じ,発電する.反対に, 可動子に電流を供給することで磁束が発生して固定子に取り付けられた永久磁石と磁極同士の吸引・反発力によ り制御力を発生する. Fig.1 のような WEC では,一般的にフロートの同調点付近以外の波周期をもつ波が卓越す る場合,発電電力量が低下する. そのため,広範囲な波浪条件下で効率よく発電するために制御が行われる²⁻⁵⁾.



Fig. 1 Schematic view and principle of the wave energy converter with a linear generator.

これまでに提案された制御法は WEC の力学モデルに基づいた制御法のため,モデル化が難しい発電機構によ る機械摩擦やコギング力により,実現象と力学モデルの誤差が生じる場合,発電電力量が低下する可能性がある. 著者らの先行研究では,モデル同定のための基礎試験を実施し,制御モデルや制御パラメータの再調整を行った ⁵⁾. 模型の場合,基礎試験によりモデル同定が可能だが,実機ではモデル同定は容易でない.また,WEC の経年 劣化により,さらに誤差が大きくなることで,発電電力量の低下および制御の誤作動などが起きることが想定さ れる.

そこで、力学モデルを用いない強化学習が注目されており、WECの制御への適用が検討されている⁶⁸.本報 告では、WECに強化学習の一種である Deep Q-network (DQN)を用いた制御の概要を紹介する⁸⁾.不規則波中の シミュレーション上で、DQN で経験から発電電力量が向上するように制御を学習させ、平均電力および電力変動 を既存の制御法を用いた場合と比較した.

2. 強化学習

強化学習の概要を Fig.2 に示す. ある状態sでエージェントが行動aを行ったときに報酬rが得られる課題を考える. 強化学習では,実際に状態sでエージェントが行動aを行い,得られた報酬rから,累積報酬を最大となるような連続行動を何らかの方策で獲得する. 設計者は行動と報酬のみを設定するので,モデルフリーな学習と呼ばれる. 本研究の場合,エージェントは発電機で,環境は波と浮体運動であり,電力量に応じた報酬を得る課題となる.



Fig. 2 Framework of the reinforcement learning⁸⁾.

強化学習の一つである Q-learning では行動価値関数Q(s, a)と呼ぶ関数で,現在の状態での全行動を評価し,取 るべき行動を決定する. Q-learning ではQ(s, a)を実際に経験に基づいて式(1)で更新する.

 $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max [Q(s',a')] - Q(s,a)]$

(1)

ここで、*a*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影響するかを調整する. *max*[*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. 式(1)中の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. TD 誤差を小さくするように*Q*(*s*,*a*)を更新する. TD 誤差が0となると、現在の状態で各行動を取ったとき、得られる報酬と遷移先での報酬期待値が予測できるようになる. 学習した*Q*(*s*,*a*)が大きい行動を選択し続けることで、累積報酬を最大化する. しかし、この手法は、状態や行動が連続値で定義され、その次元も非常に大きい環境だと、すべての状態と行動に対して定義される*Q*(*s*,*a*)を保存するための領域が膨大となってしまう. そこで、行動価値関数*Q*(*s*,*a*)をニューラルネットワークにより Q-network として関数近似する. ただし、この近似によりパラメータが増大すると、そのままでは勾配消失や過学習など上手く学習できない問題が生じる の. ニューラルネットワークによる関数近似と Experience replay や Fixed target Q-network, Dueling DQN などテクニックを使うことで学習を可能とした 9,10 . この手法を DQN と呼ぶ 9,10 . 学習効率を向上させるテクニックについては多くの書籍で解説されているので、詳細は参考文献を参照されたい 6,10 . DQN を波力発電装置に適用して制御パラメータを逐次更新する.

3. 運動シミュレーション

3.1 運動シミュレーションモデル

本研究では、固定子が取り付けられたスパー部は固定された状態とし、フロートの上下揺れのみと仮定する.計算対象としたリニア式 WEC 模型の座標系およびフロートのメッシュ分割モデルを Fig. 3 に、主要目を Table 1 に示す.将来的に検証が可能であることを念頭に置いて水槽模型サイズを対象とした.時間領域でのフロートの上下運動の運動方程式を式(2)~(4)に示す.変位や出力制限などの機械的制約を考慮していない.



Fig. 3 Schematic view of the WEC model and the mesh model of the float for calculation⁸⁾.

Part	Item	Value
Float	Outer diameter	0.3500 m
	Inner diameter	0.1140 m
	Depth	0.3000 m
	Water-plane area	0.0860 m
	Draft	0.1500 m
	Natural Period	1.00 s
	Mass	12.90 kg
	Added mass	8.00 kg
Spar	Diameter	0.1000 m
PMLG	Thrust coefficient	37.93 N/A
	Winding resistance	2.115 Ω

Table 1 Principal particulars of the WEC model.

$$(m_f + m_{\infty}) \ddot{z}(t) + \int_0^t k_{33}(t - \tau) \dot{z}(\tau) d\tau + c_{33} z(t) = f_{ext} + f_{PTO}$$

$$(2)$$

$$k_{33}(\tau) = \frac{2}{\pi} \int_0^\infty N(\omega) \cos \omega \tau \, d\omega$$

$$(3)$$

$$f_{ext} = \sum_{i=1}^M E_1(\omega_i) \sqrt{2S(\omega_i) \Delta \omega} \sin (\omega_i t + \varphi_i)$$

$$(4)$$

c t

ここで、上付きドットは時間微分、 k_{33} は造波減衰力係数 $N(\omega)$ を用いて式(3)で表されるメモリー影響関数、 f_{ext} は波強制力である. f_{ext} は $E_1(\omega)$ と $S(\omega)$ で表される M 個の規則波成分を重ね合わせることで得られる. フロート の造波減衰力,付加質量,波強制力は3次元パネル法の計算コード WAMIT により計算した¹²⁾.計算に用いたメッ シュモデルの総パネル数は約2000である.スパーとの干渉流体力は考慮していない. Fig.4 にフロートの造波減 衰力係数と波強制力係数の計算結果を示す.



リニア発電機の制御力および瞬時電力は式(5)~(8)で得られる3).制御力はフロート速度と変位に比例する力の 和で与えられ、制御パラメータ Cg, Kgを変化させて制御を行う.

$$f_{PTO} = -C_g \dot{z}(t) - K_g z(t)$$
⁽⁵⁾

$$P(t) = f_{PTO} \dot{z} \left(t \right) \tag{6}$$

$$P_c(t) = R \left(\frac{f_{PTO}}{K_t}\right)^2 \tag{7}$$

$$P_n(t) = P(t) - P_c(t) \tag{8}$$

想定するリニア発電機は電機子内部にコイルユニットを持ち,永久磁石が積層されたシャフトが電機子を貫通 している.シャフトの永久磁石によって生じる磁束とコイルに流れる電流との電磁気的作用によって推力を生み 出す.発電機側のインピーダンスを制御することで,電流および電圧を操作して任意の推力を発生させる.推力 から電流指令値を決定し,電流指令値から PI 制御により電圧指令値を決定し,PWM 制御により電圧を制御する ^{13,14}.

3.2 計算条件

不規則波中のシミュレーションを実施し、平均電力を評価した.不規則波の波条件は H_s =0.1 m, T_s =1.0~2.4 s とした.エージェントの初期状態から終端状態まで移行するまでの過程を1 Episode と呼び、本論文では1 Episode 内の評価時間は T_s ×120,刻み時間は0.01 s で 100 波以上を含む時間とした.1 つの有義波周期につき 3500 Episode の計算を行った.

波スペクトルは式(9)~(14)の JONSWAP 型波スペクトル¹⁵⁾を用いた.ここで, *f*は成分波の周波数, *c*はスペクトルの形状パラメータで 1.0 とした.不規則波を構成する成分波の位相差は Episode 毎に乱数で与えて, 異なる時系列を持つ波変位データを作成した.

$$S(\omega) = \frac{1}{2\pi} S(f) \tag{9}$$

$$S(f) = aH_s^2 T_p^{-4} f^{-5} exp[-1.25(T_p f)^{-4}] c^b$$
⁽¹⁰⁾

$$a = \frac{0.0624}{0.23 + 0.0336c - 0.185(1.9 + c)^{-1}}$$
(11)

$$b = exp - \frac{\left(T_p f - 1\right)^2}{2\sigma^2}$$
(12)

$$T_p = 1.05T_s \tag{13}$$

$$\sigma = \begin{cases} 0.07 & \text{if } f \le f_p \\ 0.09 & \text{if } f > f_p \end{cases}$$
(14)

4. 制御方法

DQN による波力発電の制御と学習のフローチャートを Fig.5 に示す.まず,制御周期毎に,現在の変位z,速度 z, C_g, K_g ,波変位 ζ を取得し,後述する 5×100×5 次元ベクトルとして状態sを作成する.状態sを入力としてニュー ラルネットワークで近似した行動価値関数を計算し,行動選択をするが,その前に,状態空間の探索を行うかど うかを ε -greedy を用いて決定する¹⁰. ε -greedy では値が 0 から 1 の間で乱数を発生させ,乱数が ϵ 以下であるとラ ンダム行動をとり,状態空間の探索を行う. 乱数が ϵ より大きい場合はニューラルネットワークで近似した行動価 値関数に基づいた行動選択を取る.探索行動を取る確率を表す ε は 0.4 から試行回数を重ねることに 0.01 まで減 少する.後述する行動テーブルの中から行動価値関数 Q が最大となる C_g, K_g の変化量($\Delta C_g, \Delta K_g$)を決定し,式 (15), (16)で更新する.

$$C_g \leftarrow C_g + \Delta C_g \tag{15}$$

$$K_g \leftarrow K_g + \Delta K_g \tag{16}$$



Fig. 5 Flowchart of the Deep Q-network for the WEC control[®].

状態sで選択した行動aをすると、報酬関数に基づいてrが得られ、行動後の変位z 速度ż、 C_g 、 K_g 波変位 ζ を表す状態s'に遷移する. これらの情報をメモリーに記憶する. この手順を制御周期である 0.1s 毎に繰り返し、十分な情報をメモリーに保存すると、Experience replay に基づいて、一旦経験した状態、行動、報酬、遷移先をメモリーに蓄積し、メモリーからランダムに取り出しバッチ学習する¹⁰. ニューラルネットワークの重みの更新量が十分小さくなれば、学習が完了となり、重みを更新しないで、行動価値関数*Q*を用いて制御を行う.

Fig. 6 に使用したニューラルネットワーク構造と DQN で定義する行動テーブル ($\Delta C_g, \Delta K_g$), 各層でのカーネ ルサイズ,ストライド,活性化関数を示す. Fig.6 中の Input が各層での入力データのベクトルサイズ,Output が 出力のベクトルサイズ,カーネルサイズは畳み込みの大きさ,ストライドはカーネルの移動距離を表す.



Fig. 6 Architectures of the neural network model for the DQN⁸⁾.

z, ż, C_g, K_g, ζ (5 変数) をサンプリング周期 0.1 s で 10 s 間のデータ 100 点を 1 セットとすると,入力データは 過去 5 セット分を合わせた 5×100×5 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 (Conv2D) を経て, Flatten で 1 次元ベクトルに変換し, 2 層の全結合型ニューラルネットワーク (Dense)の後, Dueling DQN の結合層 (Lambda)のあと, Q を出力する. 活性化関数 f(u) は,隠れ層には Rectified Linear Unit (ReLU),出力層 では線形関数を用いた⁹.ニューラルネットワークの最適化手法には Adam を使用した¹⁰. 学習のパラメータは $\alpha = 0.0003, \gamma = 0.95$ とした⁸.

報酬の重み付けはできなくなるが、報酬は変位の極大値から現在までに得た電力量から、式(17)、(18)のように 4 段階に固定化することで学習が進みやすくなる¹⁰.

$$E_{sum} = \sum_{i=1}^{m} P_n^{(i)} \Delta T$$

$$r = \begin{cases}
-1.0 & \text{if } E_{sum}/E_a < 0 \\
-0.2 & \text{if } 0 \le E_{sum}/E_a \le 0.1 \\
0.3 & \text{if } 0.1 < E_{sum}/E_a < 0.5 \\
1.0 & \text{if } E_{sum}/E_a \ge 0.5
\end{cases}$$
(17)
(17)

ただし, *m* は変位のゼロアップクロス点間の最大値から現在までのステップ数(1 ステップ ΔT =0.1 s), E_a はスケーリング係数である. 模型スケールのリニア発電機の定格出力 10 W を想定し, E_a は 10 とした.

 C_g , K_g の初期値は 5 Ns/m, -100 N/m とし,正常に Episode が終了した場合の初期値は前 Episode 終了時の値で スタートする. 初期 50 Episode はランダム行動を選択し, Experience replay 用にメモリーへ保存する.メモリー内 から 2048 個のサンプルを取り出し,バッチ学習する.また,シミュレーション中に E_{sum} が-30 を下回ると,失敗 と判断し, Episode を終了し,次の Episode に移る.

比較のために、いくつか提案されている制御法のうち Approximate Complex-Conjugate Control with Considering Copper Loss (ACL)を用いた³. ACL では発電機での損失である銅損を考慮して発電電力量を最大化する手法であり、入射する波が正弦波の場合、ACL による制御で得られる発電電力量は最大となる. ACL の K_g, C_g は式(19)か

ら(22)で決定する.入射波の周波数に応じて一意に決まるため,不規則波の場合,入射波の角周波数を有義波周期から求めた角周波数に置き換えて計算し,一定値とした³⁾.

$$K_{g} = \frac{N(\omega) + 2\delta \left(N(\omega)^{2} + X^{2}\right)}{4\delta^{2} \left(N(\omega)^{2} + X^{2}\right) + 4\delta N(\omega) + 1}$$
(19)

$$C_g = \frac{\omega X}{4\delta^2 \left(N(\omega)^2 + X^2\right) + 4\delta N(\omega) + 1} \tag{20}$$

$$X = \omega \left(m_f + m_a \right) - \frac{c_{33}}{\omega}$$

$$\delta = \frac{K_t^2}{\omega}$$
(21)
(22)

$$\delta = \frac{-r}{R}$$

5. 計算結果

Fig.7 は T_s =1.2, 1.4 s での Episode 毎の平均電力を示す. Fig.7 中の橙線と青線は T_s = 1.2, 1.4 sの Episode 毎の結果, 黒の実線および破線は T_s = 1.2, 1.4 sの 50 Episode の平均値を示す. Episode を重ねるごとに学習していき, 50 Episode 毎の平均電力が向上している. DQN を用いることで, 強化学習で適切な制御パラメータを選択可能となった.



Fig. 7 Averaged power in irregular waves (Hs=0.1 m, Ts=1.2 and 1.4 s) per episode during training⁸⁾.

1 Episode の評価時間は約 120~336 s で、平均電力評価に十分な長さでない、十分学習が進んだ 3450 Episode 後、 1 Episode での P_n の時間平均値を、50 Episode で平均化した値を平均電力とした。 波周期ごとの平均電力を Fig. 8 に示す. エラーバーは 50 Episode 中の平均電力の標準偏差、Fig. 8 中の数字は ACL の平均電力に対する DQN の 平均電力の比を表す. ACL も同様に 50 Episode の平均電力である。 T_s =1.0 s では DQN の平均電力は ACL より大 きいが、 T_s =1.2 s 以上では DQN の平均電力の方がわずかに小さい. ACL と比べて約 96~110%であった. 50 Episode の標準偏差は ACL と同程度である。Fig. 7 では Episode 毎に変動が見られるが、制御によるものでなく、不規則 波による変動といえる.



Fig. 8 Comparison between the ACL and the DQN for averaged output power in irregular waves. The error bars indicate standard deviation. The percentages show the ratio of the averaged output power in the DQN to the output power in the ACL[®].

波力発電装置では制御によって平均発電電力が大きくなるとともに、電力変動は小さいことが望ましい. Fig.9 に瞬時電力の標準偏差, Fig. 10 にピーク値/平均値 (Peak-to-Average Power Ratio: PAPR) を示す. 両図とも 50 Episode 間の平均値である. Fig.9 および Fig. 10 のエラーバーは標準誤差を表し, Fig. 9 および Fig. 10 の数字は ACL に対 する DQN の値の比である. どの波周期でも ACL より DQN の標準偏差は小さい. PAPR については最も平均電力が高い T_s =1.6 s で約 30%減少し, T_s が長くなると, ACL との差は小さくなる. T_s =1.0 s, 2.4 s では DQN の PAPR は ACL より大きいが, Fig. 8 で示した平均電力が大きい T_s =1.2 s~2.0 s では, DQN の PAPR は ACL より小さい. よって, ACL に比べて DQN のほうが電力変動は小さい.



Fig. 9 Standard deviation of the instantaneous output power in irregular waves. The error bars indicate standard errors. The percentages show the ratio of the standard deviation in the DQN to the standard deviation in the ACL⁸.



Fig. 10 Peak-to-average power ratio (PAPR) in irregular waves. The error bars indicate standard errors. The percentages show the ratio of the PAPR in the DQN to the PAPR in the ACL⁸⁾.

6. まとめ

発電機構にリニア発電機を採用したポイントアブソーバー型波力発電装置を対象として、強化学習の1つである Deep Q-network (DQN)を用いた制御を紹介した. DQN は ACL と比べて発電性能はわずかに劣るが、制御のための力学モデルが必要ない.また、ACL と比べて瞬時電力の標準偏差やピーク値と平均値との比が小さい.

謝 辞

本研究はJSPS科研費20K14976の助成を受けたものです。関係各位に深く感謝申し上げます.

参考文献

- 国立研究開発法人新エネルギー・産業技術総合開発機構: 平成 26 年度~平成 29 年度成果報告書 海洋エネ ルギー技術研究開発 次世代海洋エネルギー発電技術開発 リニア式波力発電、2018000000866 (2019).
- 2) Falnes, J., 2002, Ocean Waves and Oscillating System, Cambridge University Press, Cambridge.
- De La Villa Jaén, A., García-Santana, A., and Montoya-Andrade, D.E., 2014, Maximizing Output Power of Linear Generators for Wave Energy Conversion, International Transactions on Electrical Energy Systems, Vol. 24, No. 6, pp. 875-890.
- Hals, J., Falnes, J., and Moan, T., 2019, Constrained optimal control of a heaving buoy wave-energy converter, Journal of Offshore Mechanics and Arctic Engineering, Vol.133, No.1, 011401.
- 5) 梅田隼,後藤博樹,藤原敏文,谷口友基,井上俊司:リニア式波力発電装置のモデル予測制御に関する研究, 日本船舶海洋工学会論文集,第28号(2018), pp.27-36.
- 三上貞芳,皆川雅章ほか:強化学習,森北出版(2000).
- Anderlini, E., Forehand, D. I., Bannon, E., and Abusara, M., 2017, Control of a Realistic Wave Energy Converter Model Using Least-Squares Policy Iteration, IEEE Transactions on Sustainable Energy, Vol.8, Issue 4, pp. 1618-1628.
- 8) 梅田 隼,藤原 敏文:強化学習によるリニア式波力発電装置の電力量最大化,日本船舶海洋工学会論文集, 第31号 (2020), pp. 229-238.
- 9) Mnih, V., et al., 2015, Human-level control through deep reinforcement learning, Nature 518.7540 529.
- 10) 牧野貴樹, 澁谷長史ほか:これからの強化学習, 森北出版(2016).
- Wang, Z., Schaul, T., Hessel, M., van Hasselt, H., Lanctot, M., and De Freitas, N., 2015, Dueling Network Architectures for Deep Reinforcement Learning, arXiv preprint arXiv:1511.06581.
- Lee, C.-H., 1995, WAMIT Theory Manual., Report No. 95-2, Dept. of Ocean Engineering, Massachusetts Institute of Technology.
- 13) 紙屋大輝,後藤博樹,一ノ倉理:波力発電用リニア発電機の制御に関する検討,日本磁気学会論文特集号, 第1巻,第1号(2017), pp.57-60.
- 14) 谷口友基,藤原敏文,梅田隼,二村正,片山徹:波力発電装置の陸上試験装置の開発と実時間最適制御法の 検証,日本船舶海洋工学会論文集,第32号(2020), pp.99-108.
- 15) 合田良実:数値シミュレーションによる波浪の標準スペクトルと統計的性質,海岸工学講演会論文集, 第34巻(1987), pp. 131-135.
- 16) AKingma, Diederik P., and Ba, J., 2014, Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980.