

鉄道システムの省エネルギー実現に向けた 回生電力の充放電制御

Charge Control of Regenerative Power for Energy-Saving Railway Systems

吉田 賢央*¹ 荒井 幸代*¹
Yasuhiro Yoshida Sachiyo Arai

*¹千葉大学大学院融合理工学府 都市環境システム

Department of Urban Environment System, Graduate School of Science and Engineering, Chiba University

Recently, utilization of regenerated power from the brake operation has drawn attention to help for energy conservation of railway systems. In this paper, we introduce reinforcement learning with actor-critic to acquire the appropriate rules that decide the amount of charge/discharge so that the fluctuation of the SOC (State of Charge) can be suppressed. In the previous researches, the control rules are hand-crafted based on human empirical heuristics that has limitations on suppressing electricity supply-demand dynamics in the railway system.

We show some empirical results of both previous research and our proposed one, and find that the generated rules via reinforcement learning show the better performance than the ones by the previous researches.

1. 本研究の背景と目的

鉄道システムにおける省エネルギー実現のために、列車減速時に発生する回生電力の有効利用が注目されている。本研究では、回生電力の有効利用法として蓄電池を導入し、その充放電制御による鉄道システム内の電力需給バランスを実現する。具体的には蓄電池の充電率を抑えた制御則の獲得を目的とする。蓄電池の充電率を一定にかつ小さく保つことができれば、廉価な蓄電池設置で済むことや電気の劣化も抑えることができる。

従来の人間の経験則に基づくルールベース制御は、鉄道システム内の電力需給ダイナミクスに対して限界がある。そこで本研究では、強化学習を適用し、蓄電池充電率の安定化を目的とした、蓄電池の充放電制御アルゴリズムを提案する。

$$H = R - (A + B) \quad (1)$$

(1) 式を構成する需給電力は、列車の加速に必要な需要電力 A [kW]、回生ブレーキによる供給電力 R [kW] (回生電力)、蓄電池の充放電電流 I [A] から計算される蓄電池の需給電力 B [kW] の三つである。なお、本研究では蓄電池がエージェントであり、充放電電流 I を制御対象とする。また、蓄電池の充電率 C [%] は蓄電池の需給電力 B によって変動する。

充電率 C の安定化には理想範囲として下限 $C_{\min}$ 、上限 $C_{\max}$ を設定する。充電率 C は (2) 式、(3) 式にしたがって変動し、下限 $C_{\min}$ と上限 $C_{\max}$ を超えない充放電電流 I を制御する。蓄電量と蓄電池容量を E [kWh]、 $E_{\max}$ とし、制御間隔を Δt [s] とする。

2. 問題設定

2.1 電力需給モデル

図 1 に本研究で対象とする鉄道システム内の電力需給モデルを示す。列車、変電所、蓄電池は架線を介して電力需給する。(1) 式に鉄道システム内の電力需給バランス H [kW] を定義する。

$$C_t = \frac{E_t + B_t \Delta t}{E_{\max}} \cdot 100 \quad (2)$$

$$E_t = \frac{E_{\max} C_{t-1}}{100} \quad (3)$$

2.2 既存手法

蓄電池への入力情報は架線電圧 (Overhead power line Voltage) OV [V] と充電率 C である。架線電圧 OV は電力需給バランス H が正のとき上昇し、負のときに降下する性質をもつ。

既存手法では上記の性質を利用したルールベースの制御則を用い、架線電圧 OV を入力し、充放電電流 I を出力する。図 2 に既存手法で用いる制御則を示す。この制御則により電力需給バランス H の安定化を考慮している。ただし図 2 の制御則は充電率 C の安定化を考慮していないルールを設定している。

現実には電力需給バランス H と充電率 C の両面を考慮したルールは存在する。しかし、ルールベース制御では独立した多変数を入力する場合、ハンドコードですべてのルールを明確に決定することは困難である。また、ハンドコードによる制御則は人間の経験則に基づいているため、最適保証をもたない。

これまでにも充電率 C を考慮した充放電制御の研究 [Satake 14] がされてきた。文献 [Satake 14] の研究では、鉄道システム内の消費電力と回生電力が均衡すると推測した場合

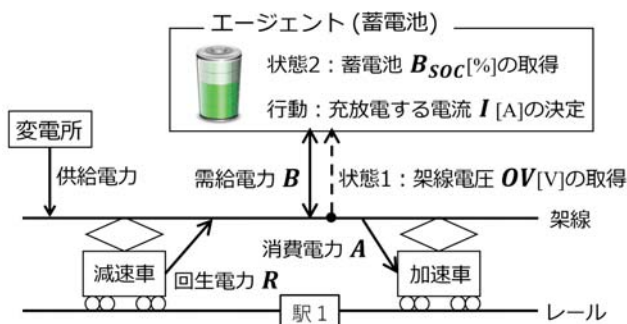


図 1: 鉄道システム内の電力需給モデル

連絡先: 吉田 賢央, 千葉大学大学院融合理工学府 都市環境システム, 千葉市稲毛区弥生町 1-33, aesa2088@chiba-u.jp

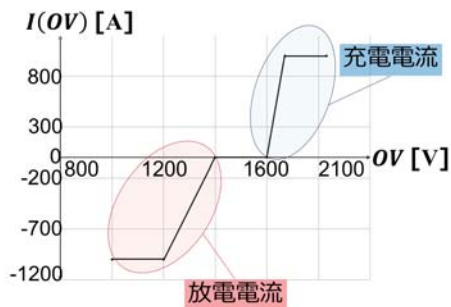


図 2: 既存手法の制御則例

Algorithm 1 Actor-Critic を用いた学習方法

状態 s , 正規分布 $\mathcal{N}(\mu_s, \sigma_s)$ を初期化

$t = 0$ に初期化

while $t \leq t_{\text{terminal}}$ **do**

状態 s_t を観測

Actor で正規分布 (μ_s, σ_s) に従った方策 π から

行動 a_t を出力

次状態 s_{t+1} を観測

報酬 r_{t+1} を観測

// 正規分布 $\mathcal{N}(\mu_s, \sigma_s)$ の更新

$$V(s_t) \leftarrow V(s_t) + \alpha(r_{t+1} + \gamma V(s_{t+1}) - V(s_t)) \quad (4)$$

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \quad (5)$$

if $\delta_t \geq 0$ **then**

$$\mu_s \leftarrow \mu + \beta_\mu \delta_t (a_t - \mu) \quad (6)$$

end if

$$\omega_s \leftarrow \omega_s + \beta_\omega \delta_t ((a_t - \mu)^2 - \sigma^2)(1 - \sigma) \quad (7)$$

$$\sigma_s = \frac{1}{1 + \exp(-\omega_s)} \quad (8)$$

end while

に、充電率 C から初期値に向けて充放電する制御を取り入れている。これにより、充電率 C についても考慮した充放電制御を実現している。しかし、文献 [Satake 14] の制御則は鉄道システム内の消費電力と回生電力の均衡が極めて少ない環境において、充電率の初期値に向かう制御がとられず、充電率 C の考慮はできない。

3. 提案法

本研究では強化学習の一つである TD(0) 法 [Sutton 98] を用いて蓄電池の充放電制御則を獲得する方法を提案する。強化学習は、環境モデルが未知で、ハンドコードによるルールが記述できない場合でも報酬による試行錯誤によって最適な制御則の獲得が期待できる。

TD(0) 法はある時刻 t の状態 s_t の価値 V_{s_t} を (4) 式によって更新する。 r_{t+1} は次の時刻 $t+1$ における状態 s_{t+1} への遷移時に得られる報酬値であり、エージェントは r_t が最大とな

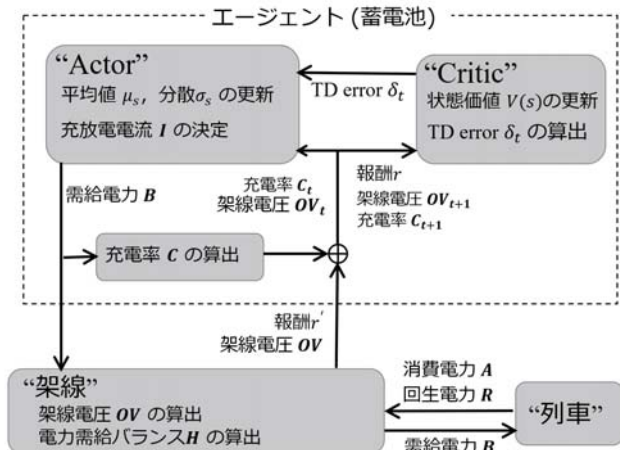


図 3: 鉄道システムにおける学習エージェントモデルの全体像

Algorithm 2 蓄電池の充放電制御

エージェントにおける状態 s , 方策 $\pi(\mu_s, \sigma_s)$ を初期化

$t = 0$ に初期化

while $t \leq t_{\text{terminal}}$ **do**

// 充放電制御

架線電圧 OV , 充電率 C を観測

方策 π にしたがって充放電電流 I を決定

(2) 式にしたがって充電率 C を更新

// 価値 V , 方策 π の更新

価値 $V(s)$, 方策 $\pi(\mu_s, \sigma_s)$ の更新

end while

る行動を選択する。なお、 α, γ はそれぞれ学習率、割引率を示す。ここで、TD(0) 法で扱える状態と行動はともに離散表現であるのに対し、対象問題では、「状態入力として、架線電圧 OV と蓄電池充電率 C 」、「行動出力として、蓄電池の充放電電流 I 」といずれも連続量である。そこで、TD(0) 法への状態入力を離散化し、TD(0) 法で出力される離散行動ではなく、電力の制御量として連続行動を出力する。その出力法に Actor-Critic [Kimura 00] を用いる。

3.1 Actor-Critic を用いた離散行動の出力

Algorithm1 に Actor-Critic を用いた学習手順を示す。1 エピソードの最終ステップを t_{terminal} とする。Actor-Critic は行動選択部 Actor と状態評価部 Critic から構成されており、提案法では各状態 s での行動 a_s を正規分布 $\mathcal{N}(\mu_s, \sigma_s)$ に基づいて確率的な連続行動で出力する。 μ_s と σ_s の値は TD(0) 法の更新式の差分である TD 誤差 (TD error) δ_t によって更新する。提案法では $\delta_t < 0$ のときには σ_s だけを更新し、 μ_s については更新しない更新方法 [Kagaya 13] を用いる。なお、 ω_s はシグモイド関数を用いて σ_s が 0 以下をとらないための内部変数、 β_μ, β_ω は方策割引率である。また、学習中における局所解への収束と発散を防止するため、 μ_s, σ_s の値域はそれぞれ $-4.0 \leq \mu_s \leq 4.0, 0.0 \leq \sigma_s \leq 100.0$ とする。

3.2 蓄電池の充放電制御への適応

図 3 に鉄道システム内における電力および、エージェントが扱う情報の流れを示し、Algorithm2 に強化学習エージェントによる充放電制御アルゴリズムを示す。まず、時刻 t における鉄道システム内の需給電力 A_t, R_t, B_t によって架線電圧 OV および、充電率 C が毎時刻変化する。次に、充放電制御は刻

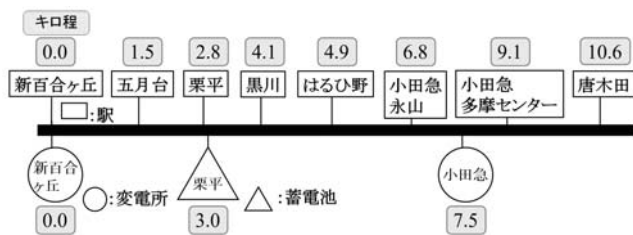


図 4: 対象線区の路線図

み時間 Δt 毎に行われ、状態入力として $s = (OV, C)$ を観測する。最後に、この状態入力からエージェントは充放電制御を行い、行動出力として充放電電流 I を出力する。なお、提案法では Algorithm1 によって出力される行動 a_t の値域が充放電電流 I の値域と異なるため、最終的に出力された値に適当な係数を乗算し、充放電電流 I として出力する。

エージェントには充電率 C と電力需給バランス H の安定化に関する報酬 r が (9) 式、(10) 式に基づいて与えられる。

$$r'_t = \begin{cases} B_{\max} - |H_t| & \text{if } -B_{\max} \leq R - A \leq B_{\max} \\ -B_t & \text{if } R - A < -B_{\max} \\ B_t & \text{if } B_{\max} < R - A \end{cases} \quad (9)$$

$$r_t = \begin{cases} r'_t & \text{if } C_{\min} < C_{t+1} < C_{\max} \\ r'_{\min} & \text{otherwise} \end{cases} \quad (10)$$

(9) 式は電力需給バランス H の安定化に関する報酬であり、蓄電池の充放電制御によって電力需給バランス H を抑えた電力を報酬 r' として算出する。ただし、蓄電池による需給電力 B には上限 $B_{\max}$ 、下限 $B_{\min}$ が存在するため、上限 $B_{\max}$ または $B_{\min}$ を鉄道システム内の電力 $R - A$ が超える際には需給電力 B または $-B$ を報酬 r' とする。また、充電率 C が理想範囲を超えたときに報酬 r' の最小値 $r'_{\min}$ を報酬 r とする。

報酬を与えられたエージェントは観測した状態の価値 $V(s)$ と平均値 μ_s 、分散 σ_s を (4) 式、(6) 式、(8) 式にしたがって更新する。

以上の処理を繰り返すことによって充電率 C を考慮した制御則を獲得する。獲得した制御則を鉄道システムに適用し、鉄道システム内の電力需給バランス H を安定化する。

4. 提案法の有効性に関する実験

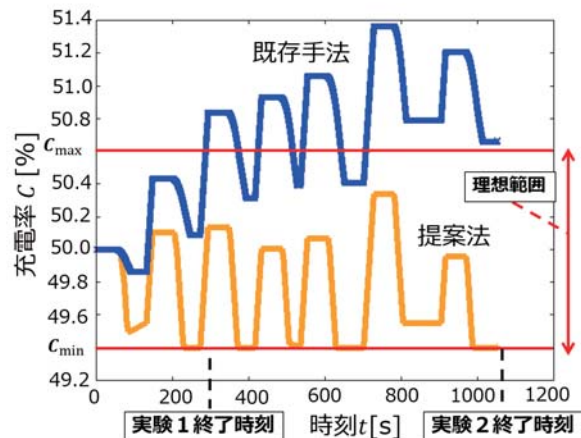
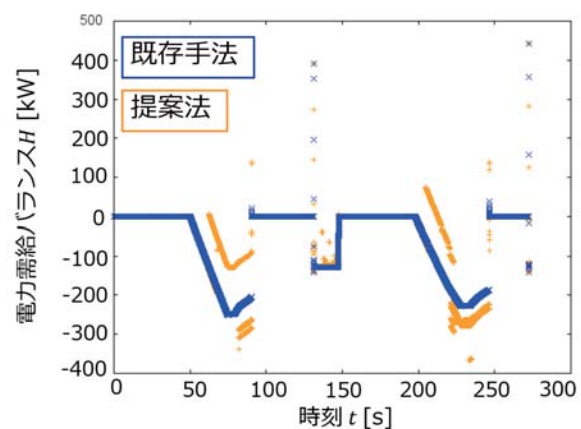
提案法による蓄電池の充放電制御の有効性を二つの実験によって確かめる。

4.1 実験設定

図 4 に計算機実験で対象とする鉄道システムに関して駅、変電所、蓄電池の位置関係を示す路線図を示す。実験 1 では図 4 における新百合ヶ丘駅から栗平駅の 3 駅、実験 2 では対象線区すべての 8 駅を 1 台の列車が走行する。

刻み時間 $\Delta t = 0.001$ とし、時刻 t 毎の充放電電流 I ($-200 < I < 200$) を決定する。状態入力は架線電圧 OV ($1200 < OV < 1820$)、充電率 C ($0 < C < 100$) である。本実験では初期架線電圧 $OV_0 = 1620$ 、蓄電池の初期充電率 $C_0 = 50$ とする。報酬 r は 3.2 節にしたがって毎時刻与える。強化学習のパラメータは学習率 $\alpha = 0.001$ 、割引率 $\gamma = 0.5$ 、方策割引率 $\beta_\mu = \beta_\sigma = 0.001$ とする。

充電率 C の安定化に向けた理想範囲 $C_{\min} \leq C \leq C_{\max}$ を、実験 1 の環境で既存手法を用いた場合の充電率変動幅 ($49.4 \leq$

図 5: 充電率 C の時間推移図 6: 実験 1 における電力需給バランス H の時間推移

$C \leq 50.5$) とする。充電率変動幅は初期充電率 C_0 を基準としたときに、最も変動した変動量を上限 $C_{\max}$ 、下限 $C_{\min}$ とした範囲である。

4.2 実験結果および考察

■蓄電池充電率 C の安定化

図 5 に実験 2 における充電率 C の時間推移を示す。縦軸は充電率 C 、横軸は時刻を示している。既存手法では蓄電池充電率の変動が、実験 2 の終了時刻までに理想範囲を超過しているのに対し、提案法では理想範囲内に抑えられていることがわかる。したがって、提案法により充電率 C の安定化を考慮した制御則の獲得ができたと考えられる。

■電力需給バランス H の安定化

図 6、表 1 に電力需給バランス H の安定化に関する実験結果をまとめる。図 6 は実験 1 における電力需給バランス H の時間推移を示し、縦軸は電力需給バランス H 、横軸が時刻 t を示している。図 6 より $t = 60$ から $t = 90$ 、および $t = 210$ から $t = 250$ のときで提案法と既存手法に電力需給バランス H の差異が確認できる。この結果から提案法の性能を評価することは困難だと考えられる。

ここで、表 1 から総電力量、および最大最小値を比較する。電力需給バランス H の総電力量について、実験 1 および実験 2 の両方で提案法は既存手法に対して総電力量を削減できた。

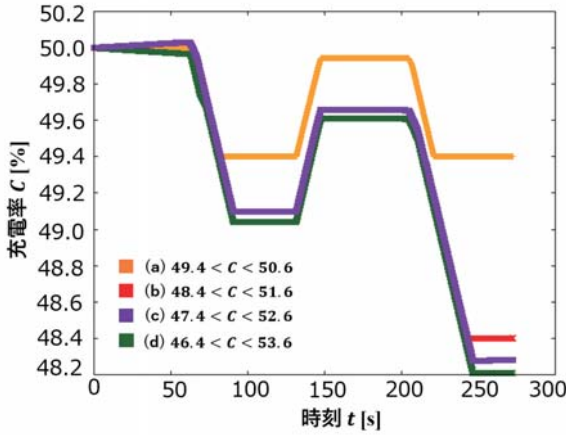
最大最小値について、最大値は提案法と既存手法に大きな差はない。最小値は提案法の方が低く、電力需給バランス H

表 1: 提案法と既存手法の電力需給バランス H

実験	制御手法	電力需給バランス H		
		総電力量	最大値	最小値
実験 1 (3 駅)	既存手法	45.65	441.97	-249.51
	提案法	39.01	441.97	-349.24
実験 2 (8 駅)	既存手法	158.72	441.97	-249.51
	提案法	157.11	441.98	-396.08

表 2: 理想範囲ごとの電力需給バランス H

実験	理想範囲		電力需給バランス H		
	$C_{\min}$	$C_{\max}$	総電力量	最大値	最小値
(a)	49.4	50.6	39.01	441.97	-349.24
(b)	48.4	51.6	32.31	441.97	-327.9
(c)	47.4	52.6	31.41	441.97	-304.2
(d)	46.4	53.6	31.39	441.97	-317.17

図 7: 理想範囲ごとの充電率 C の時間推移

の一時的なピークは提案法の方が大きかった。これは一時的なピークを抑える報酬設計ではなかったためであり、新たな報酬設計が必要である。

以上から提案法は、既存手法と比較して同等以上の制御結果を示した。とくに、充電率 C の安定化については既存手法よりも充電率の変動が安定した制御結果であった。よって、提案法では既存手法のルールベース制御では対応できない制御則の獲得が期待でき、有効であると考えられる。

5. 提案法の適応性に関する実験

5.1 実験設定

実験 1 で設定した環境から充電率 C の理想範囲を上限下限 1% ずつ最大 3% まで緩和させる。本実験では (a) $49.4 \leq C \leq 50.6$, (b) $48.4 \leq C \leq 51.6$, (c) $47.4 \leq C \leq 52.6$, (d) $46.4 \leq C \leq 53.6$ の 4 通りを充電率 C の理想範囲として実験する。

強化学習のパラメータは学習率 $\alpha = 0.001$, 割引率 $\gamma = 0.5$, 方策割引率 $\beta_\mu = \beta_\sigma = 0.001$ とする。実験 1 と同様に図 4 における新百合ヶ丘駅から栗平駅を 1 台の列車が走行する。上記の環境で提案法による蓄電池の充放電制御を行い、設定する充電率 C の理想範囲に対して適応性の有無を確認する。

5.2 実験結果および考察

■蓄電池充電率 C の安定化

図 7 に充電率 C の時間変動を示す。縦軸は充電率 C , 横軸は時刻 t を示している。設定したすべての理想範囲において、各理想範囲内で制御ができていた。(a), (b) の場合は列車走行中に各理想範囲の下限 $C_{\min}$ まで変動しているのに対し, (c), (d) では理想範囲の下限 $C_{\min}$ には至らなかった。これは、充電率 C の理想範囲に関係なく電力需給バランス H の安定化に向けて最適方策 π が獲得できたためだと考えられる。

■電力需給バランス H の安定化

表 2 に電力需給バランス H の安定化に関する実験結果をまとめる。充電率 C の緩和が大きいほど総電力量の削減も大き

いことがわかる。ただし, (d) では (c) のときと比べてほとんど同じ値を示した。これは、前述したとおり充電率 C の理想範囲に関係なく電力需給バランス H の安定化に向けて最適方策 π が獲得できたためだと考えられる。

最大値は理想範囲の緩和にかかわらず、同じ値を示した。最小値については (d) のときを除いて、理想範囲を緩和をするほど一時的なピークを抑えた。

理想範囲が (d) のときは (c) のときと総電力量がほとんど同じ値であり, (c) のときが提案法の中で電力需給バランスの一時的なピークを最も抑えている。ただし、既存手法と比べると総電力量は大きく抑えられているが、一時的なピークは抑えられていない。4.2 節での考察と同様に、一時的なピークを考慮した報酬設計でないためだと考えられる。

本実験で充電率 C の理想範囲を緩和した際に各理想範囲に対応した制御則獲得ができた。以上より提案法による充放電制御の充電率 C の安定化に関する適応性が確認できたと考えられる。

6. 結論および今後の課題

本研究では、鉄道システムの省エネルギーの実現に向けて電力需給バランスを実現し、充電率を安定化する蓄電池の充放電制御アルゴリズムとして制御則の獲得に強化学習を用いた。計算機実験では、提案法が充電率と電力需給バランスの安定化の両面から既存手法と比較し有効であり、設定する充電率の理想範囲に対して適応性があることを確かめた。

一方、現状では提案法によって電力需給バランスの一時的なピークを抑えることはできていない。これは、ピークを抑えることを考慮した報酬設計をしていないからである。そこで、さらなる電力需給バランスの安定化に向け、一時的なピークを抑える充放電制御アルゴリズムの設計が今後の課題である。

また、本研究ではエージェントが蓄電池だけのシングルエージェントである。今後は列車の運転行動も制御対象に加えたマルチエージェント環境への拡張を検討する。

参考文献

- [Satake 14] 佐竹信彦, 野木雅之 and 保科俊一郎, “鉄道向け回生電力蓄電システム,” 東芝レビュー 69.8, pp.52-55, (2014)
- [Sutton 98] R. S. Sutton and A. G. Barto, “Reinforcement learn-ing: An introduction,” MIT press Cambridge, Vol.1, No.1(1998)
- [Kimura 00] 木村元 and 小林重信, “Actor に適正度の履歴を用いた Actor-Critic アルゴリズム: 不完全な Value-Function のもとでの強化学習,” 人工知能学会誌, Vol. 15, No. 2, pp. 267-275 (2000)
- [Kagaya 13] 加賀谷駿 and 荒井幸代, “太陽光発電の変動抑制に向けた強化学習エージェントモデル,” 電子情報通信学会論文誌 D, 96.12, pp.3000-3008, (2013)