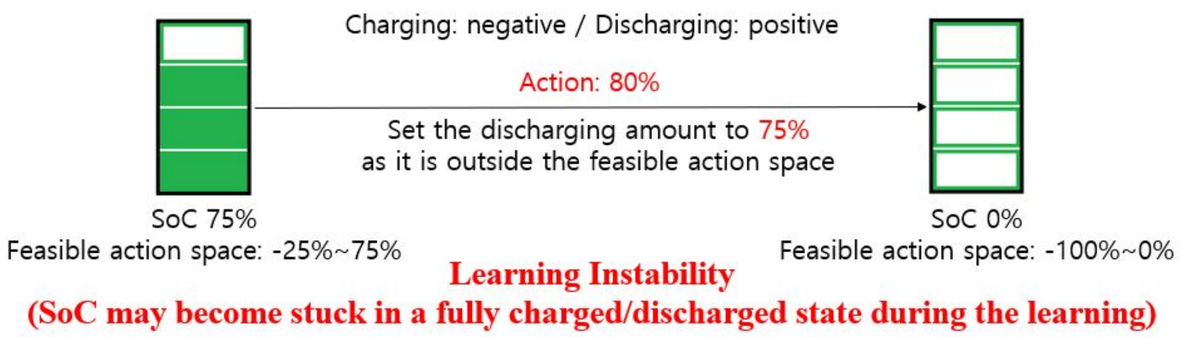
Time-Varying Constraint-Aware Reinforcement Learning for Energy Storage Control

Paper ID: 17

Jaeik Jeong, Tai-Yeon Ku, Wan-Ki Park
Energy ICT Research Section, Electronics and Telecommunications Research Institute (ETRI)

Introduction

- Energy storage devices are crucial for reducing the impact of climate change by efficiently storing excess energy from renewables and releasing it during peak demand, thus decreasing reliance on fossil fuels [1].
- In recent times, reinforcement learning has gained prominence over traditional optimization methods for energy storage control by enabling dynamic adaptation for energy storage systems [2,3].
- However, time-varying feasible charge-discharge range based on state of charge (SoC) variability limits the conventional reinforcement learning.
- We propose a reinforcement learning approach that takes into account the time-varying feasible charge-discharge range. An additional objective function was introduced for learning the feasible action range.
- This actively promotes the utilization of energy storage by preventing them from getting stuck in suboptimal states, such as continuous full charging or discharging.



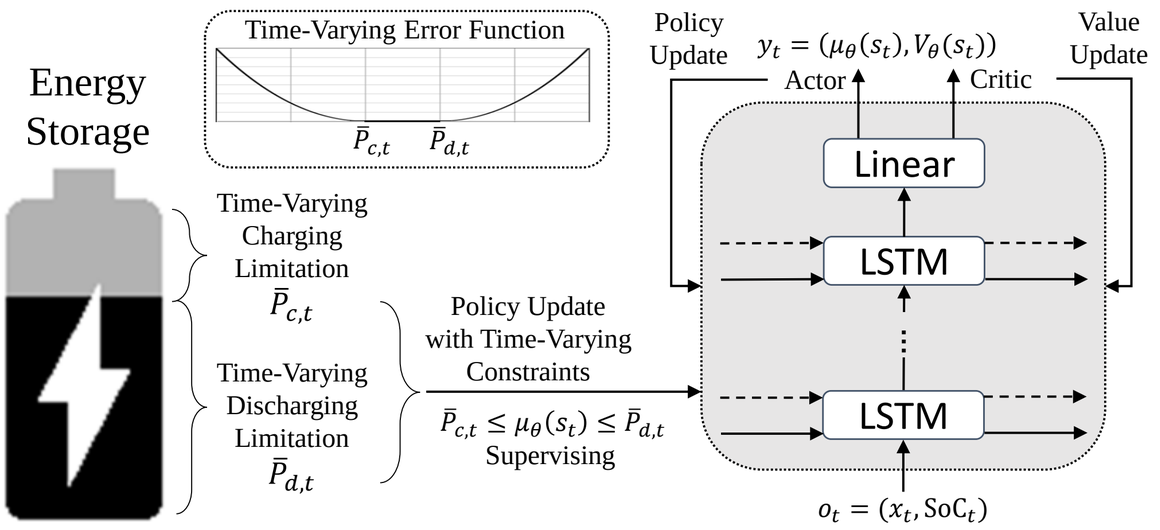
Methods

- We adopt the proximal policy optimization (PPO) algorithm with long-short term memory (LSTM) model [4].
- With the parameters θ and the state s_t, the output consists of the actor's output $\mu_\theta(s_t)$, and the critic's output $V_\theta(s_t)$.
- During the training phase, action a_t is sampled from the Gaussian policy with mean $\mu_\theta(s_t)$, and during the actual testing phase, $\mu_\theta(s_t)$ serves as the action a_t.
- An additional objective function was introduced for learning the feasible action range for each time period, supplementing the objectives of training the actor for policy learning and the critic for value learning.
- Let the charging limitation as $\bar{P}_{c,t}$ and the discharging limitation as $\bar{P}_{d,t}$. The proposed supervising objective function is as follows (charging actions are negative, and discharging actions are positive):

$$L^{PPO}_{\text{supervising}}(\theta) = \min\left(\mu_\theta(s_t) - \bar{P}_{c,t}, 0\right)^2 + \min\left(\bar{P}_{d,t} - \mu_\theta(s_t), 0\right)^2. \quad (1)$$

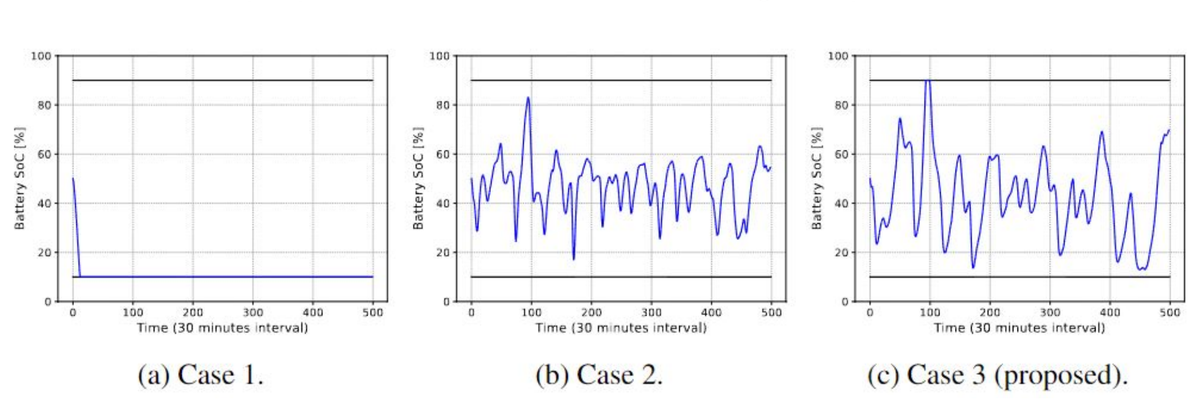
- We finally obtain our main objective, which is minimized at each iteration:

$$L^{PPO}(\theta) = L^{PPO}_{\text{actor}}(\theta) + C_1 L^{PPO}_{\text{critic}}(\theta) + C_2 L^{PPO}_{\text{supervising}}(\theta), \quad (2)$$



Results

- We demonstrated the effectiveness of the proposed approach through energy arbitrage experiments based on actual energy price data [5].
- We take the first 2000 data points which are sampled every 30 minutes and split the dataset into training set (1000 data points), validation set (500 data points), and test set (500 data points) in chronological order.
- At time slot $t = 0$, we set $\text{SoC}_t = 0.5$, i.e., the half stored energy. We set the minimum and maximum values of the SoC to 0.1 and 0.9, respectively, in order to prevent battery degradation.
- We normalize the price data between 0 and 1 by the maximum price \$190.81/MWh. We simulate the proposed method using 100MWh battery with the degradation cost of \$10/MWh.
- Case 1 employed a conventional continuous reinforcement learning approach, excluding the equation (1). It shows a scenario where all the initially stored energy is discharged (sold out) and no further actions are taken. This suggests a failure in learning to manage the costs associated with charging in energy arbitrage, resulting in suboptimal behavior.
- Case 2 incorporated the equation (1) into the reward function. This approach adds negative rewards if the output actions fall outside the feasible action space, rather than explicitly learning the range of output actions. It demonstrates reasonable utilization of energy arbitrage.
- Case 3 designates the proposed model. It engages in more active energy arbitrage. Introducing equation (1) as a negative reward makes the agent conservative towards reaching states of complete charge or discharge, leading to reduced utilization of the energy storage.
- The table below presents the 30-minute average of charging cost, discharging revenue, degradation cost, and total profit for the three cases. It is evident that the proposed Case 3 achieves the highest profit.



(a) Case 1. (b) Case 2. (c) Case 3 (proposed).

Experiment results (30-minutes averaged)

	Metric			
	Charging cost (\$)	Disharging revenue (\$)	Degradation cost (\$)	Total profit (\$)
Case 1	-0.000	4.858	-0.080	**4.779**
Case 2	-40.039	53.245	-1.691	**11.514**
Case 3 (proposed)	-37.493	54.085	-1.714	**14.879**

Conclusion

- We introduce a continuous reinforcement learning approach for energy storage control that considers the dynamically changing feasible charge-discharge range.
- An additional objective function has been incorporated to learn the feasible action range for each time period. This helps prevent the energy storage from getting stuck in states of complete charge or discharge.
- Furthermore, the results indicate that supervising the output actions into the feasible action range is more effective in enhancing energy storage utilization than imposing negative rewards when the output actions deviate from the feasible action range.
- In future research, we will explore combining offline reinforcement learning or multi-agent reinforcement learning to investigate methods for learning a more optimized policy stably.

KEY REFERENCES

[1] Mathew Aneke and Meihong Wang. Energy storage technologies and real life applications–a state of the art review. Applied Energy, 179:350–377, 2016.

[2] Mostafa Rezaeimozafar, Maeve Duffy, Rory FD Monaghan, and Enda Barrett. A hybrid heuristic-reinforcement learning-based real-time control model for residential behind-the-meter PV-battery systems. Applied Energy, 355:122244, 2024.

[3] Jaeik Jeong, Seung Wan Kim, and Hongseok Kim. Deep reinforcement learning based real-time renewable energy bidding with battery control. IEEE Transactions on Energy Markets, Policy and Regulation, 2023.

[4] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017.

[5] The changing price of wholesale UK electricity over more than a decade. https://www.ice.org.uk/knowledge-and-resources/briefing-sheet/the-changing-price-of-wholesale-uk-electricity, 2017.

Electronics and Telecommunications Research Institute

MORE INFORMATION

Jaeik Jeong
Electronics and Telecommunications Research Institute (ETRI)
Energy ICT Research Section

jaeik1210@etri.re.kr

https://jaeik-jeong.github.io/blog/