A Reinforcement Learning Formulation for Optimal Control of Energy Storage Devices

February 17, 2024

1 Introduction

This document elaborates on a reinforcement learning framework designed for the optimal control of energy storage devices. The framework's objective is to minimize electricity costs, considering fluctuating demand and prices, while adhering to the physical and operational constraints of the energy storage system.

2 Reinforcement Learning Framework

The framework is defined as a discrete-time Markov Decision Process (MDP), with elements $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where:

- $\mathcal{S}$ denotes the state space,
- $\mathcal{A}$ represents the action space,
- P is the state transition probability function,
- R is the reward function,
- γ is the discount factor for future rewards.

2.1 State Space

The state $s \in \mathcal{S}$ is defined as a vector including the demand d_t, the state of charge SOC_t, the electricity price p_t, and time-encoded features. These features capture cyclical patterns in electricity use and pricing, which are significant for decision-making.

2.2 Action Space

The action $a \in \mathcal{A}$ dictates the adjustment in the battery's charge level, with discretized options such as $\{-0.1, 0, 0.1\}$, representing the energy to be either withdrawn from or injected into the battery.

2.3 Reward Function

The reward function aims to reduce electricity expenditure, expressed as:

$$R(s, a) = -\left(p_t \cdot (d_t + a \cdot E)\right),$$

where E denotes the battery's energy capacity.

2.4 Transition Function

The transition function $P(s'|s, a)$ models the dynamics of how the current state and action chosen by the policy lead to the next state. It incorporates both the physical process of charging or discharging the battery and the progression of time, reflecting changes in demand, price, and temporal features. For instance:

$$SOC_{t+1} = SOC_t + a \cdot E,$$

subject to battery capacity limits and efficiency losses, capturing how the battery's state of charge evolves.

2.5 Objective

The reinforcement learning agent seeks a policy $\pi^* : \mathcal{S} \to \mathcal{A}$ that maximizes the expected cumulative reward over a defined time horizon:

$$\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R(s_t, \pi(s_t))\right].$$

3 Constraints and Considerations

3.1 Battery Constraints

The battery's physical and operational constraints are critical to the model:

- **State of Charge Limits:** SOC_t must remain within the minimum and maximum bounds to prevent overcharging or deep discharging, which can harm the battery's lifespan and efficiency.
- **Rate Limits:** The rate of charging or discharging is constrained to prevent excessive stress on the battery, ensuring operational safety and longevity.

3.2 Operational Constraints

Operational constraints also include ensuring the system's actions are feasible within the grid's limitations and comply with regulatory requirements, such as demand response events or peak load management.

3.3 Environmental and Economic Considerations

While optimizing for cost, the model can also consider environmental impacts, such as reducing reliance on fossil fuels during peak demand times by smartly managing the storage device's charge and discharge cycles.

4 Conclusion

This reinforcement learning framework provides a structured approach to optimizing energy storage device control. By modeling the system's dynamics and constraints, it aims to achieve cost-effective, efficient, and sustainable energy management.