Data-Driven Micro Grid Optimization using Reinforcement Learning

Trenton Griffiths, Jeff Wasden, and Nicholas S. Flann

Utah State University

Abstract

The rapid deployment and optimal operation of Micro-grids is an essential step in increasing the adoption of solar power, improving grid resilience and expanding electrical power availability in developing countries. Micro-grids integrate local energy generation and storage, local loads and run in isolation or with a grid connection. Given time-of-day grid energy pricing, predictions of solar energy production and knowledge of upcoming loads, it is possible to determine the optimal operating policy. This paper presents a system that combines historical weather and usage data with neural-network prediction and reinforcement learning methods to determine this policy. night. Results show the learned policy significantly reduces operating costs.

Data-Driven Micro Grid Optimization using Reinforcement Learning

Introduction

Reinforcement learning is a powerful and yet universal method that learns the best actions to take in all states of a domain. To identify the optimal operating policy, the applicant need only specify the set of actions possible in each state, the resulting new state, and the benefit or loss of the action, called the immediate reward.

Before we consider the micro-grid, consider a generic domain of well defined states and actions. We will assume that the state is observable and that a deterministic model exists of the domain that returns the next state given an action. Under these simplified assumptions, the problem can be represented as a value function learning problem and solved using a form of dynamic programming.

Given S_t, the state at t, then $S_{t+1} = A(S_t, a)$, where a is the action and Do executes the action, producing the next state. For a given state S_t, the set of actions available are $A(S_t)$.

Then let us define $V(S_t)$ as the total reward for the domain being in state S_t. Optimal operation will always take actions such that the states occupied maximize total reward. Then $V(S_t)$ may be defined as:

$$V(S_t) = \max_{a \in A(S_t)} R(S_t, a) + \gamma V(Do(S_t, a)) \tag{1}$$

Where $R(S_t, a)$ is the immediate reward of being in state S_t and taking action a, and $0 < \gamma \leq 1.0$ is the discount rate.

The optimal policy always selects the action that maximizes future total reward. Then the optimal policy is defined:

$$\Pi(S_t) = \arg \max_{a \in A(S_t)} R_t(S_t, a) + \gamma V(Do(S_t, a))$$

Reinforcement learning is a method that solves for V and Π, given the model. In this

case, the model is a specification of a representation of S, and a definition of $A()$, $Do()$ and $R()$. Additionally, there may be some states for which there are no actions available and the total reward is known, these states are referred to as terminal with $t = T$.

V is complied into a look-up table, mapping the current state to a number. Consider this table of all possible states, each with a V value. The value iteration method first assigns values to terminal states, then repeatedly applies Equation 1 as an update rule until all the values in the table are at a near fixed point. The order of the updates could be made in a specific way, such as from the terminal states backwards, or asynchronously based on states that are experienced during operation.

Micro-Grid Application

The state of the micro-grid is the energy held in the battery at any given time.

$S = \langle Batt_t \rangle$, where $0 \leq Batt \leq Batt_{MAX}$. Let us pick discrete times, such as every 15 minutes, and integer values of $Batt$, representing some quantization of Kwh.

The actions $A(\langle t, Batt_t \rangle)$ are all possible charge/discharge events ΔE_i such that $0 \leq Batt_{t+1} = Batt_t + \Delta E_i \leq Batt_{MAX}$

The time period of the domain is limited to T hours and represented at 15 minute intervals, providing $0 \leq t \leq T * 4$ distinct time iterations.

The environmental state in which the micro-grid policy operates is based on the charge in the battery. Using the state, the action is chosen to charge or discharge the battery at any given time, t.

State update functions

$Batt_{t+1} \rightarrow Batt_t + \Delta E_t$

Reward functions

$R_t(S_t, a) = R_{Buy/Sell}(S_t, a) + batteryLoss(a, t) - batteryPercentagePenalty(S_t)$

$batteryPercentagePenalty(S_t) = 20 - percentCharged$

$R_{Buy/Sell}(S_t, a) = (solar[t] + a - demand[t]) * buySell[t]$

$$batteryLoss(a, t) = -|a^2|$$

Methods

The method used to optimize the micro-grid is a value iteration algorithm (VIA). The VIA will calculate the optimal reward for each state the Micro-grid can enter and then choose an action that will advance the state to the next, most beneficial state. The VIA will do this by using the energy generated by the solar panels, the current load on the Micro-grid, and the current state of the battery, for all possible states.

At the end of a given time period the battery percentage can be in any of its predefined states, for example a battery with 14 000 watt hour capacity could be broken into ten, 1 400 watt hour bins. Using the final state of the battery, the VIA calculates the value of the energy stored in the battery at that time. Then the VIA calculates the value of being in the state previous, until each time step for the period is filled out for every possible state.

By calculating an intermediate reward, based on the purchased or sale of electricity plus the wear on the battery, the learning algorithm will be able to calculate the ideal times to purchase electricity from the main power-grid. It will also be able to decide when to sell the stored energy back to the main power grid, and when to use the energy stored in the battery to accommodate the load on the Micro-grid.

Conclusion

By using a reinforcement learning algorithm to optimize a Micro-grid, the system reduced its operating cost by 75% on average. This was accomplished using simulated weather and load data, where all of the information was known. Further work that could be implemented on this project would be to implement a neural-network to predict the solar generation, based on weather data, and the load on the Micro-grid. However, by implementing these tools the VIA would no longer be deterministic, and therefore not accurate; changes would need to be made to make the model probabilistic. By being able

to predict weather patterns and systems loads, the savings of the system would be further increased as it would then compensate for any irregular scenario, or new situation that the Micro-grid was placed in.

References

Reinforcement learning for microgrid energy management, https://www.sciencedirect.com/science/article/pii/S036054421300481

Reinforcement Learning-based Energy Trading for Microgrids, https://arxiv.org/pdf/1801.06285.pdf

Reinforcement Learning: An Introduction, http://www.incompleteideas.net/book/the-book.html