Reinforcement Learning Methods for Energy Microgrids

Ben Moore
06274155
Stanford University
Aeronautics & Astronautics

Sander Tonkens
06274594
Stanford University
Mechanical Engineering

Bruis van Vlijmen
06276015
Stanford University
Mechanical Engineering

Abstract

In this paper different model free reinforcement algorithms are explored to optimize energy storage distribution in small/micro-grids. The microgrid features renewable energy sources, a large-scale PV system and a wind farm, and two storage capacities, a pumped hydro storage and a lithium-ion battery, to meet the demand of a small town. The problem of optimally storing/supplying every excess/deficit in the grid is modeled as a sequential decision making under uncertainty problem where, at every time-step, the uncertainty comes from the lack of knowledge about future electricity consumption and renewable energy production sources. Different reinforcement learning algorithms have been implemented to extract knowledge from the current storage levels to decide the optimal action to take. Two generalization methods are used to determine the action to take for non-visited states. Q-learning combined with a Neural Network approximation yields the best results for an agent unaware of the local energy inbalance, however a heuristic approach which is aware of the energy excess/deficit and attempts to store/supply this excess/deficit randomly over the different storage capacities yields better results.

Problem Description

An electricity microgrid is a group of interconnected loads and distributed energy resources within clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid. A microgrid can connect and disconnect from the grid to enable it to operate in both grid-connected or island-mode, as defined by the U.S. Department of Energy Microgrid Exchange Group [1]. The energy system consists of local energy generation, local energy consumption and storage capacities [2]. In this paper, we consider two local electricity production sources, a large-scale photovoltaic (PV) system and a wind farm. In addition, we consider different types of storage devices to address fluctuations of local electricity production and local demand. One of the main challenges consists of finding a storage strategy capable of handling uncertainties related to future electricity production and consumption, as stated by Franois-Lavet et al. [3]. This paper analyzes a problem similar to Franois-Lavet et al., but on a larger scale (city-size compared to residential setting) and introduces different solving methods.

For this paper, we consider a remote village of 50.000 inhabitants, which mostly relies on a large-scale PV system and a wind farm to meet its energy demand. To meet fluctuations in the demand, it relies on the following storage capacities: a high-capacity lithium-ion battery (LIB) (similar to Tesla's mega-battery in South Australia [4]) and a pumped hydro storage (PH). As mentioned above, the grid can be enabled to operate in grid-connected mode to meet high electricity demand. In this paper we investigate several reinforcement learning techniques and approximation methods which determine how the storage capacities should interact (i.e. which action to take) with the microgrid, under what circumstances.

Prior Work

Previous work regarding microgrids has mostly focused on single consumer problems. Different reinforcement learning techniques have been proposed. Kuznetsova et al. implemented Q-learning with 2-step lookahead for battery scheduling. Predictions of available wind power, through a Markov chain model, feed the reinforcement learning algorithm for selecting the optimal battery scheduling actions [5]. Dimeas and Hatziargyriou implement multi-agent Q-learning to control residential-sized microgrids [6]. However, most of the research that combines reinforcement learning with microgrids, focuses on smart grids and how to optimize suppliers and consumers behavior, for instance Foruzan et al. [7]. Instead of focusing on single-consumer demand and one storage option, we focus on a larger scale remote village with different storage options. Instead of only considering a broad storage strategy, the agent needs to decide when to make use of which storage type with the goal to maximize its reward.

Power Balance

An electricity grid can never have a net excess/deficit and as such supply always meets demand. In a microgrid a deficit in local energy production gets compensated by supplying electricity from the grid. In case of local surplus energy, the microgrid evacuates the excess by wasting through heat release into the environment.

The physical system has to satisfy the following equations at each time step (P denotes that it is supplied to the system, whilst d denotes it is being pulled from the system):

$$\delta_t = P_t{}^W + P_t{}^{PV} - d_t{}^d \tag{1}$$

$$\delta_t = d_t{}^w - P_t{}^{LIB} - P_t{}^{PH} - P_t{}^{EG} \tag{2}$$

where:

$P_t{}^W$ = Electricity generated locally by wind power
$P_t{}^{PV}$ = Electricity generated locally by PV system
$d_t{}^d$ = Local consumption
$P_t{}^{LIB}$ = Electricity transferred out of (+) or into (-) LIB
$P_t{}^{PH}$ = Electricity transferred out of (+) or into (-) PHS
$P_t{}^{EG}$ = Electricity supplied by regular grid
$d_t{}^w$ = Excess electricity supplied

State Space

The state space, $\mathcal{S}$, consists of the storage operation state (i.e. the amount of energy in the storage devices) and the time of day.

Storage The amount of energy in the lithium-ion battery is denoted by $s_t{}^{LIB}\,[Wh]$ and the amount of energy in the pumped hydro storage is denoted by $s_t{}^{PH}\,[Wh]$. In addition, we introduce $x^i[Wh]$ as the capacity, $\eta_{in}{}^i\,[\%]$ and $\eta_{out}{}^i\,[\%]$ as the charge efficiency (i.e. when storing) and discharge efficiency (i.e. when delivering) and $\zeta^i\,[\%]$ as the self-discharge rate of storage i. The Results section includes a table with the sizing of the above-mentioned variables for our purpose.

To feed our state space into the state-action matrix, we discritize $s_t{}^i$ uniformly (between 0 and x^i).

Time The time of day, t, is discretized based on the time interval of the available data.

Action Space

At every time step, an action $a_t = [a_t{}^{LIB}, a_t{}^{PH}] \subseteq \mathcal{A}_t$ is applied on the system, where $a_t{}^i$ is the amount of energy transferred out of (if positive) or into (if negative) storage i. In addition to the variables defined in the Storage section, we introduce $P_{max,out}{}^i$ and $P_{max,in}{}^i$, the power ratings in and out of storage i. The Results section includes a table with the sizing of the above-mentioned variables for our purpose.

Physical Constraints

As defined in the Action Space section, the action space is bounded. As the action space directly influences the state space, only a subset of states s_{t+1} can be visited from state s_t. The relationship between states and actions are governed by the following equations:

$$s_{t+1}{}^i = \frac{1}{1-\zeta^i}\left(s_t{}^i - \frac{a_t{}^i}{\eta_{out}{}^i}\right) \tag{3}$$

when energy is transferred out of storage i.

$$s_{t+1}{}^i = \frac{1}{1-\zeta^i}\left(s_t{}^i - \eta_{in}{}^i a_t{}^i\right) \tag{4}$$

when energy is transferred into storage i ($a_t{}^i < 0$). In addition, when transitioning state we move from time t to time $t+1$.

Reward

The reward function of the system, the local electrical grid, corresponds to the instantaneous operational costs r_t. Costs are made up of the cost of the electricity that was not supplied within the microgrid and the waste energy. The equation for r_t is as follows:

$$r_t = r^{EG} + r^{waste} = -k_t{}^{EG} \cdot P_t{}^{EG} - k_t{}^w \cdot d_t{}^w \tag{5}$$

where:

$k_t{}^{EG}$ = cost of importing electricity per Wh
$k_t{}^w$ = emission tax per Wh wasted

Data/Simulator

The time of day component of the state space is discretized based on the data available. As such, we require the data to include multiple entries per day. In addition, the observations included in the data are wind power production, solar power production, energy consumption and the cost of importing electricity from the grid. Using real data, uncertainty is encoded in our observation. From a specific dataset we proceed as follows:

1. Scale power consumption to a population of 50.000
2. Scale wind and solar power production such that their mean value is on the same order of magnitude as the mean power consumption, ensuring that the maximum power production is higher than the maximum power consumption.

Solving Methods

As a first step, a heuristic approach is taken to measure the performance of random actions during the test simulation to provide a baseline cumulative reward. Both model based and model free algorithms are then implemented to determine which algorithm produces the best policy. Given the large size of the Q matrix, the sparseness of this matrix after exploration and to further improve the policy's performance, global approximation is applied in the form of a Neural Network, Linear and Non-Linear regression.

Heuristic Approaches

As an initial baseline approach, a heuristic approach is implemented. This heuristic iterates over the test data and at each timestep, takes a random action to either store or supply from storage. After selecting an action to store or supply the heuristic then selects the amount of energy to either be transferred to or out of each storage unit randomly. The score produced by this approach serves as baseline for the solvers implemented in this paper.

Building on this heuristic, the information available to the heuristic was increased by making the demand, solar and wind production at each timestep available as part of the state space. This allows this omniscient heuristic to choose an optimal action from either store or supply at each timestep, it then takes random action to choose the amount to store or supply. The score produced by this heuristic serves as a model in which supply and store actions are optimal (i.e. it stores/supplies exactly the right amount up to the maximum power ratings) but storage energy usage is sub-optimal.

Model Based Algorithms

As the microgrid with storage capacities model represents a physical process, it deals with physical limitations. Taking these physical limitations (which limit the next states that can be accessed from the current state) into account, Model Based Reinforcement Learning can be implemented.

Maximum Likelihood Model

A Maximum Likelihood Model-Based method is proposed to leverage the fact that the model has a limited number of actions at each state and as such cannot visit every state given the current state. Therefore, it is possible to initialize $N(s, a, s')$, which counts the number of transitions from (s, a) to s', according to this knowledge. In the case of the microgrid the state space $\mathcal{S}$ is defined by three discretized physical variables: ${s_t}^{LIB}$, ${s_t}^{PH}$ and t. These variables are discretized by respectively 31 x 101 x 24 different states.

Not all states however, can be visited from each other state, due to a limit imposed by the maximum storage in- or efflux. The maximum rates of electricity transferred out and into storage i are ${P_{max,out}}^i$ and ${P_{max,in}}^i$ respectively. This means that from the t to $t+1$, the storage state variables can only change from ${s_t}^i$ to the next state bounded by replacing ${a_t}^i$ by ${P_{max,out}}^i$ in equation 3 and ${P_{max,in}}^i$ in equation 4.

Based on this knowledge, the $N(s, a, s')$ tensor can be initialized, assigning non-zero values to the $N(s, a, s')$ elements which satisfy the physical bounds imposed by the equations above.

In addition, the time component of the state space changes deterministically, further limiting $N(s, a's')$.

In this case, the state space $\mathcal{S}$ is discretized over 75144 states and an action space $\mathcal{A}$ discretized over 132 actions, the size of $N(s, a, s')$ is $|\mathcal{S}^2| \cdot |\mathcal{A}|$. It was attempted to implement this model based approach however, the large dimensional $N(s, a, s')$ tensor put too much strain on the memory capacity of the computing resources available. Maximum Likelihood Model-Based Reinforcement Learning is more suitable for lower dimensional problems.

Model Free Algorithms

A number of model free methods were implemented to find the optimal model free algorithm for this project. In this paper, the state action value matrix is referred to as the Q matrix.

Q-Learning

Q-Learning was initially implemented independently, producing a highly sparse Q matrix, which did not produce an effective policy, therefore global approximation solutions and eligibility traces were explored and implemented, reducing the sparseness of the Q matrix. In Q-Learning, the Q matrix is updated via the following equation [8].

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma \max_a Q(\acute{s}, a) - Q(s, a)) \quad (6)$$

Due to the sparsity of the resulting Q-matrix after 1 iteration, the data is iterated multiple times to achieve a better filling of the Q-matrix. As the exploration strategy is random it will not lead to over-fitting. Increasing the number of iterations over the dataset improved the accuracy of global approximation, as evidenced by the improved total reward over the test dataset. The number of iterations and the learning rate were tuned via the total reward achieved by the algorithm on the test dataset.

Sarsa-Lambda Learning

In an effort to reduce processing time and explore the performance of other model free algorithms Sarsa-Lambda was implemented. The Sarsa-Lambda algorithm also produces a highly sparse Q matrix, therefore the same techniques applied to Q-learning to address this problem were applied. Similar to the process in Q-Learning, the number of iterations over the training data, the learning rate and decay parameter, were tuned for the Sarsa-Lambda algorithm. The Q matrix is updated via the following equation [8].

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma Q(s', a') - Q(s, a)) \quad (7)$$

Eligibility Traces

Another method to reduce the sparseness of the Q-matrix is through an Eligibility Traces algorithm, which assigns credit to past states and actions. In the context of the energy microgrid problem as presented in the Problem Description section, this process is equivalent to assigning a larger negative reward to previous state and action pairs which were observed en route to a negative reward. This process improves the performance of both the Q-learning and Sarsa-Lambda algorithms, however it also significantly increases the processing time, limiting the number of iterations over the data.

Generalization

Linear & Non-Linear Regression

Linear regression is a global approximation method which can be implemented to predict the state-action pairs of the Q matrix which have not been observed during exploration. This is a practical method to fully populate the Q matrix, as fully populating the Q matrix through observations alone is computationally infeasible. The following equation is used to determine the weights for each of the features.

$$\theta = (x^T x)^{-1} x^T y \quad (8)$$

Where y is the vector of state action values observed during exploration and x is the vector of features used to predict the state action values. The state-action value, q, was then predicted using the weights, θ computed above in equation 8 and the feature values at that particular state and action, F.

$$q = F \cdot \theta \quad (9)$$

Features were chosen and expanded upon based on the sum of the reward collected when the optimal policy produced was tested on the test dataset. Initially only linear terms were included, however using the above measure of performance, higher order terms were included (Non-linear regression), which improved performance.

Neural Network

Linear and non-linear regression are based on the the notion of distance, hence they work better when approximating states that are in closer vicinity to the visited states. When using global approximation methods, such as perceptrons, the Q-matrix can be approximated using equation 10 below:

$$Q(s,a) = \theta^T \beta(s,a) \tag{10}$$

Here, θ represents the perceptron weights. When combining these perceptrons in a Neural Network, non-linearities can be imposed through calling upon non-linear activation functions. By learning the weights of the neurons in the network, back-propagation can train the network to learn to map $Q(s,a)$ values based on the feature vector input.

The network constructed for the smart-grid problem is a Sequential network with three hidden layers hidden layers and 7-dimensional input layer, to accommodate the 7 feature-variables.

Hidden 1 7 nodes, dense layer, ReLU activation

Hidden 2 14 nodes, dense layer, Sigmoid activation

Hidden 2 8 nodes, dense layer, ReLU activation

Figure 1 shows the Sequential layers in a schematic depiction. The network trained on the sparse Q-matrix data for 50 epochs, using *adam* optimization before globally approximating the unvisited state-action pairs in the Q-matrix.

Data

The algorithms described in the Solving Methods section are trained and tested, over 2017 and 2018 data respectively, from the Finnish grid [9]. The state-action space is explored randomly. Knowing δ_t from equation Power Balance, the exploration strategy assumes it can store/supply every excess/deficit, δ_t. The following equation defines the action taken if $\delta_t > 0$:

$$a_t{}^{PH} = \max\left[P_{max,out}{}^i, -\min\left[x_t{}^i \cdot \delta_t, P_{max,in}{}^i\right]\right] \tag{11}$$

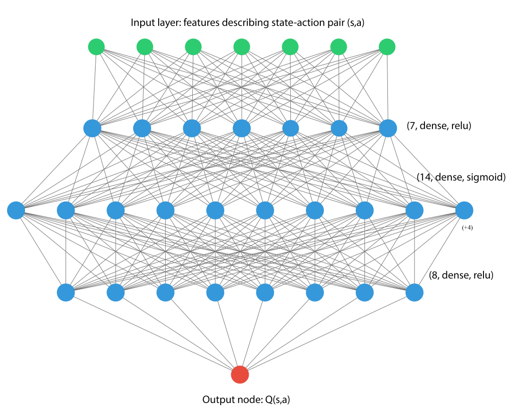

Figure 1: Neural Network schematic representation

Table 1: Storage performance data

	PH	LIB
Power rating ($P_{max}{}^i$)	100 MW	100 MW
Energy capacity (x^i)	2000 MWh	300 MWh
Storage efficiency (η^i)	80%	65%
Self-discharge rate (ζ^i)	0.01%	0.3%

with:

$$\sum_{i=1}^{n} x_t{}^i = 1 \tag{12}$$

where:

x_t^i = share of the excess/deficit that is stored in storage i.

The data is scaled as described in the Data/Simulator section. The data source is FinGrid which keeps track of electricity demand, the share of all electricity sources and the cost per Wh for the full country of Finland. The electricity consumption data is scaled down by a factor 100, to match the typical consumption of a town of 50,000 inhabitants. The wind and solar power generation are scaled such that the yearly wind power generation accounts for $\sim 80\%$ of the yearly electricity consumption and the yearly solar power generation accounts for $\sim 20\%$ of the yearly electricity consumption. The graphs below (insert reference) shows the daily fluctuation of the net surplus/deficit.

Storage Sizing

Table 1 shows the storage performance data of the pumped hydro and lithium-ion battery. The energy capacity of the storage devices are sized such that the town can be powered by the storage devices for the duration of approximately one day. This is on the same order of magnitude of energy storage devices for small islands, such as Tesla's mega-battery on Ta'u, an Island in American Samoa [10].

The power rating, charge efficiency and self-discharge rate are based on previous work on quantifying the technical features of Energy Storage Systems (ESS) [11].

The self-discharge rate is the 24-hour discharge rate, whilst it is considered that the charge efficiency and discharge efficiency are the same and satisfy:

$\eta_{in}{}^i = \eta_{out}{}^i = \sqrt{\eta^i}$

Results

The algorithms used are referred to with the following nomenclature;

- QLNN = Q-Learning and Neural Network Global Approximation
- QLTNN = Q-Learning with Traces and Neural Network Global Approximation
- SLNN = Sarsa-Lambda Learning and Neural Network Global Approximation
- SLTNN = Sarsa-Lambda Learning with Traces and Neural Network Global Approximation

Storage

The physical behaviour of the battery, defined in the Physical Constraints section, is verified from outputs during exploration over the training data. The following graphs, Figure 2 and Figure 3, demonstrate the storage units supplying or storing energy cumulatively (MWh) over a 200 hour period. The max rate of energy transfer (MW) is evident from the δ_t line, which does not exceed the max rate for both storages.

Table 2 below shows the cumulative reward function for the different solving methods with different approximation methods.

Solver

The usage of the LIB and PH storage units by each solver in response to delta over a sample period of 200 hours is shown below in Figures 4, 5, 6 and 7.

Table 2: Optimal cumulative reward

	Linear regression	Neural network
Q-learning	$-2.09 \cdot 10^6$	$-1.94 \cdot 10^6$
Sarsa-λ	–	$-2.11 \cdot 10^6$
Q-learning w/ traces	–	$-2.36 \cdot 10^6$
Sarsa-λ w/ traces	–	$-2.23 \cdot 10^6$
Heuristic baseline	$-3.41 \cdot 10^6$	
Heuristic omniscient	$-1.22 \cdot 10^6$	

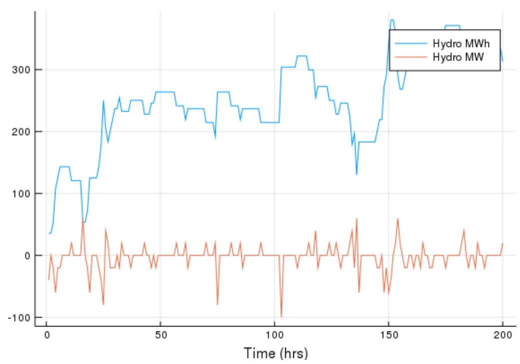

Figure 2: Hydro storage timeseries

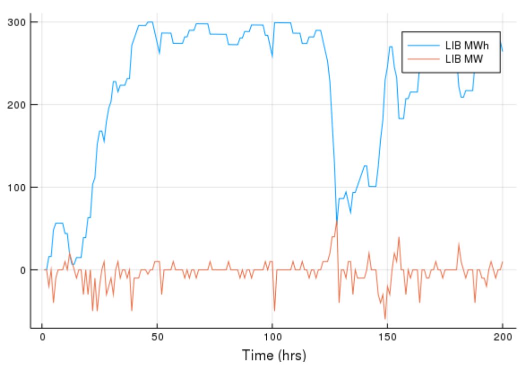

Figure 3: LIB storage timeseries

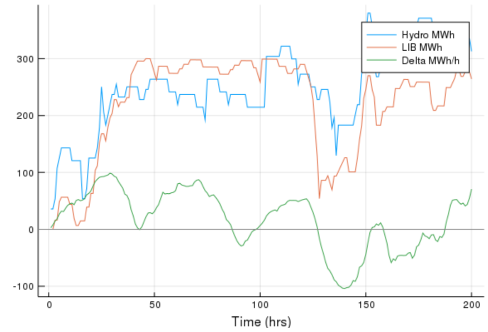

Figure 4: Storage of QLNN with Delta

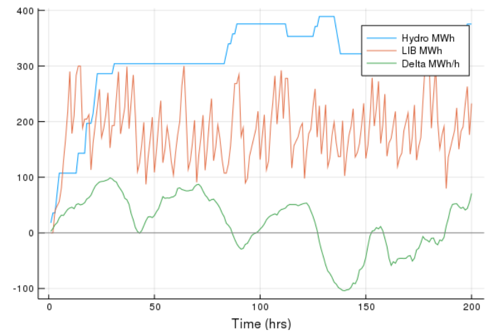

Figure 5: Storage of QLTNN with Delta

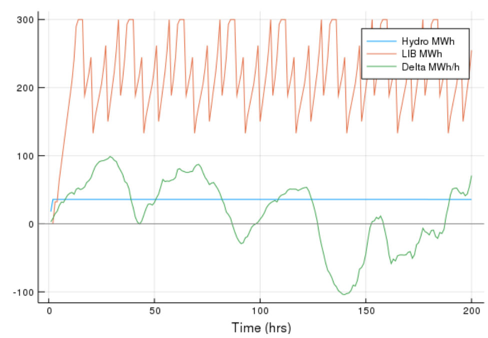

Figure 6: Storage of SLTNN with Delta

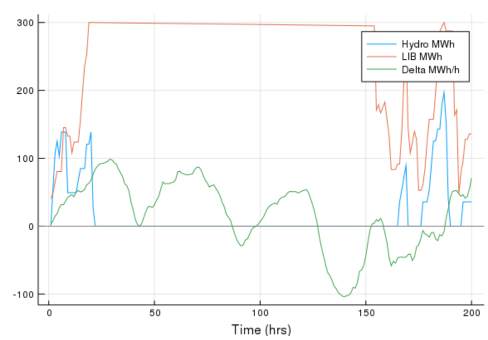

Figure 7: Storage of SLNN with Delta

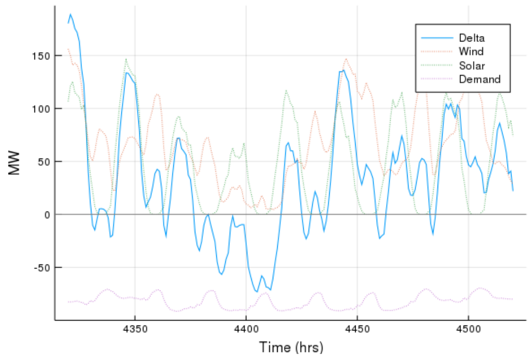

Figure 8: Energy production and consumption in July 2018

Environment

A sample period of approximately one month is presented below, depicting the; demand, ${d_t}^d$, solar energy production, ${P_t}^{PV}$, wind energy production, ${P_t}^W$ and the delta energy sum, δ_t at each timestep of one hour. The delta energy sum is calculated using equation 1. Figure 8 above shows the data for March 2018.

Discussion

In this section of the paper, the different results are evaluated and interpreted. The following results were produced using a learning rate, α of 0.6 and a decay rate, λ of 0.8. Both values used for each variable produced the best results during testing.

Storage Timeseries

To show that the storage levels of both ${s_t}^{LIB}$ and the ${s_t}^{PH}$ react directly to any energy supplying from or storing to the storage units according to the behaviour described in equations 3 and 4.

Storage Levels

As seen in figure 4, the smart grid responds to d_t deficits. These d_t deficits occur whenever the green line in the 4 drops below 0, meaning the local grid can not meet all of its demand using renewable sources. Then, in times where it has energy stored in its batteries, it will supply energy out of those storage units. This phenomenon can be seen at timestep 130 in figure 4.

Behavioural Observations

By inspection of Figure 4, it can be seen from the first 50 hours, that the QLNN solver has learned to store energy to storage when there is a surplus of renewable energy given a certain demand, in other words when δ_t is positive. It can also be seen from timestep 125 to 150 that the system has learned to supply energy from storage when there demand is greater than current renewable production, when δ_t is negative.

Similarly in Figure 5, it can be seen that the QLTNN has also learned to store energy to storage at times when δ_t is positive however it does not respond as well as the QLNN solver when there is a negative δ_t. This is demonstrated from timestep 125 to 150 in 5, when the QLTNN solver does not pull energy from storage to meet demand and thus incurs a negative reward by supplying energy from the grid.

In Figure 6, it is clear given the repeating storage amounts in LIB, that the SLNN solver has learned a storage strategy for a single day. Unfortunately, it has neglected to make use of the PH storage, which demonstrates that the solver has not explored sufficient actions and states for the PH storage unit.

The SLNN solver accesses the PH storage unit in a similar manner to the QLNN solver as shown in Figure 6, responding to a positive δ_t by storing energy into PH storage. It can be seen form Figure 6, that the solver fails to make use of the both the LIB and PH storage during timesteps 25 to 165, when δ_t is positive. The solver does perform as desired when δ_t becomes negative at timesteps 155 to 180, by supplying energy from storage. The solver could however perform better as it does not respond to the negative δ_t at timesteps 125 to 150.

Linear Regression

Linear and Non Linear regression was applied to the Q matrices produced by all model free solvers. Unfortunately, these processes could only produce valid Q matrices for Q-Learning and failed to predict valid and reasonable state-action values for the other model free algorithms. For this reason tests were not conducted on these erroneous Q matrices.

Conclusion

In this paper, different reinforcement learning techniques and approximation methods were evaluated for learning an optimal policy for an agent operating a local grid. The agent, supplying energy to a town of 50,000 inhabitants could choose between supplying or storing different amounts of energy from energy storage units. The action policies that the agent has learned through the different reinforcement learning methods, determine whether the local grid can rely on its storage units in times of an energy deficit or whether it is forced to pull expensive energy from an external source.

It appears from the results that Q-learning, combined with a Neural Network global approximation yields the best result of the attempted reinforcement learning methods.

Comparing the two different approximation methods used, the Neural Network performed better for the Q-learning then Linear Regression did for that same learning method. For the other learning strategies, approximation using Linear Regression did not yield proper results. Whereas the Neural Network yields increasingly more expensive results over Sarsa-λ, Sarsa-λ with eligibility traces and Q-learning with eligibility traces.

In conclusion, Q-learning combined with a Neural Network approximation yields the best results: $-1.94 \cdot 10^6$. Comparing this result to the reward achieved by applying the superior, omniscient heuristic which yields $-1.22 \cdot 10^6$, Q-learning method proposed is not too far off.

Future Work

Including the observations, ${P_t}^W$, ${P_t}^{PV}$ and ${d_t}^d$, in the state space will most likely lead to better decision-making. In addition, as Finland's solar production is very seasonally dependent, including the month of the year and training and testing the algorithms over multiple years might lead to better results.

It would be interesting to investigate how the trained algorithm performs in a different environment (i.e. different dataset). This would determine whether the agent takes its actions based on the fluctuations in the data (in which case training it on different data would lead to approximately the same results) or based on the actual values (in which case it could be a very bad method).

Lastly, including a storage functionality with positive reward, such as done in previous work with hydrogen storage [3], to add an element of complexity to the decisions. It would be interesting to see how tuning the previously presented algorithms would favor maximizing immediate positive reward or minimize supplying electricity from the grid in the long run, hence minimizing negative reward.

Team Member Contribution

Ben Moore: implementing RL algorithms, linear & non-linear regression for global approximation.
Bruis van Vlijmen: Neural Network for global approximation, model-based solving techniques.
Sander Tonkens: Defining state and action space, literature review, sizing simulation data
We worked together to generate and analyze results, for the discussion and the conclusion.

References

[1] "The U.S. Department of Energy's Microgrid Initiative". In: *The Electricity Journal* 25.8 (2012), pp. 84–94.

[2] Vincent Francois-Lavet et al. "Towards the Minimization of the Levelized Energy Costs of Microgrids using both Long-term and Short-term Storage Devices". In: *Smart Grid: Networking, Data Management, and Business Models* (2016), 295fffdfffdfffd319.

[3] Vincent Francois-Lavet et al. "Deep Reinforcement Learning Solutions for Energy Microgrids Management". In: (2016).

[4] "Tesla mega-battery in Australia activated". In: *BBC* (2017, December 1st).

[5] Elizaveta Kuznetsova et al. "Reinforcement learning for microgrid energy management". In: *Energy* 59 (2013), pp. 133–146.

[6] A. L. Dimeas and N. D. Hatziargyriou. "Multi-agent reinforcement learning for microgrids". In: (July 2010), pp. 1–8.

[7] E. Foruzan, L. Soh, and S. Asgarpoor. "Reinforcement Learning Approach for Optimal Distributed Energy Management in a Microgrid". In: *IEEE Transactions on Power Systems* 33.5 (Sept. 2018), pp. 5749–5758.

[8] Mykel J. Kochenderfer. *Decision Making Under Uncertainty, Theory and Application*. MIT Lincoln Laboratory, 2015.

[9] FinGrid OpenData. *Open data on the electricity market and the power system*. Electricity market and power system data retrieved from the Finnish grid, `https://data.fingrid.fi/en/`.

[10] Danielle Muoio. "Here is how Tesla uses its giant batteries to power small islands". In: *Business Insider* (2016, October 10th).

[11] Haoran Zhao et al. "Review of energy storage system for wind power integration support". In: *Applied Energy* 137 (2015), pp. 545–553.