CS 229 - Optimizing a hydrogen-electricity supply chain

Category : Reinforcement Learning

Malik Boudiaf
Department of Aeronautics and Astronautics
Stanford University
Stanford, USA
Email: mboudiaf@stanford.edu

Calvin Mc Sweeny
Department of Mechanical Engineering
Stanford University
Stanford, USA
Email: calvinmc@stanford.edu

Zoe Ghiron
Department of Aeronautics and Astronautics
Stanford University
Stanford, USA
Email: zghiron@stanford.edu

I. Introduction

Reaching the 2050 carbon-neutral goal that was set during the 2016 Paris agreement requires changes in the management of energy supply at a World-scale. While shifting towards renewable generation is part of the answer, the penetration of these new technologies requires a rearrangement of the actors of the electricity supply chain in order to ensure the security and the resilience of the grid. In the mobility sector and for some industrial sectors such as metallurgy, Hydrogen seems to be a possible contender to replace carbon-based energy vectors such as oil and gas. Complementing electricity in sectors where electrification is not cost-effective or unfeasible, hydrogen would allow a greater decarbonization of the economy as a whole.

In this project, we propose to study the optimization of a simple hydrogen/electricity supply chain. Numerous classic optimization techniques have been applied to this problem. Yet, we were not able to find any paper in the literature using machine learning techniques for this task. Hence, we propose to compare classic optimization techniques' to machine learning algorithms' performance in taking optimal energy management decisions. This comparison will be made on the basis of a real world data based simulation.

II. Problem Statement

For this project, we adopt the point of view of a sustainable energy supplier company that owns a solar farm and a wind farm. In addition, the company has decided to own storage capacity in order to take advantage of the varying electricity price on the wholesale market. But instead of just owning battery packs as most companies do, the company has bought a mix of hydrogen pressure vessels and battery packs to penetrate the mobility supply chain as well. The objective is to play on both the electricity wholesale market and the hydrogen supply chain to maximize the benefits. In addition to the storage units, the company acquired water electrolyzers to perform the electricity to hydrogen conversion internally :

In our scenario, the company interacts with a simplified environment :

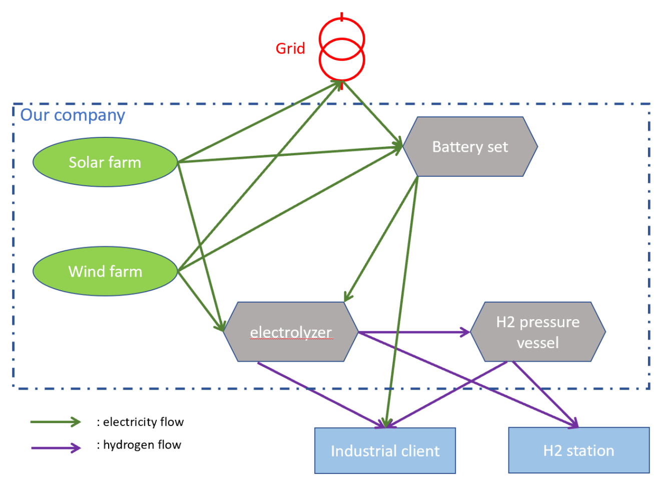

Fig. 1. Vanilla hydrogen/electricity supply chain

- *The grid :* The grid is a simplified representation of the national electrical grid. It may include household as well as other energy companies. Our company can buy and sell electricity to and from the grid
- *Industrial clients :* Refineries, ammonia plants and factories in the metal industry require a continuous supply flow of hydrogen
- *H_2 station:* The company also aims at providing hydrogen for cars and a fleet of hydrogen trains, which only require to be supplied a couple times a week

The objective of the company is to take optimal actions time step of the simulation, in the sense that it should maximize its profit over the whole simulation. Here, an action means deciding of the quantity to flow through each arrow in Fig.1.

III. Approach

The problem is formulated as an unsupervised learning problem. Indeed, even if we were able to find labeled data, where labels would here be actions taken at each time, how to be sure those actions are optimal ? Thus, why would be want to train an agent to mimic these data ? Instead, we're basically able to say that a good policy should yield a lot of money. Hence, it sounds intuitive to work in the reinforcement learning framework.

In this project, we will try to compare three different types of policies. The first most basic one, that will serve as a baseline, will consist in always taking the greediest action - the one that maximizes the immediate reward -. The second type of policy will rely on classic optimization theory, and will try to be as close as possible to what is currently implemented in many industrial models. Finally, the third type will consist of a set of models trained using different assumptions, with different RL algorithms, that are yet to be determined.

IV. DATASET

For the solar and wind farms, we used the data delivered by [4]. These data give the amount of produced power in the selected solar and wind farms in Belgium, sampled every 15 minutes, in MW. We use a normalized version of the data and scale it to the size of the farms we want to simulate. This source was very convenient because both wind and solar data were given by [4]. In addition, it is the only source we found that delivered data for specific solar farms (and not an entire region - with individual solar arrays).

For the price of electricity, we used the data supplied by [5] which gives the price of electricity $(/MWh)$ for several countries across Europe.

V. MDP FORMULATION

The main goal of our project is to determine a policy π such that in the state s, we execute the action $\pi(s)$ in order to maximize our benefit. Our problem can be formulated as a Markov Decision Process (MDP). In such setting, the agent chooses action a_t at time t based on observing state s_t. In this section, we present the setup of the MDP formulation.

A. The state space

A state must account for every distinctive feature that will have an impact on the action to be chosen. We decided to include in our state space :

- $s^{(1)}$ time of the day,
- $s^{(2)}$ the price of electricity on the market,
- $s^{(3)}$ our global electricity production (wind and solar farms)
- $s^{(4)}$ the current battery charge level,
- $s^{(5)}$ the current electrolyzer production rate,
- $s^{(6)}$ the current H_2 pressure vessel charge level.

The H_2 delivered to our clients (industrial or stations) is fixed because it relies on hourly contracts. We could make this more complex later on (for example if the demand changes with time).

This implies that we know exactly our state space. However, the state space is continuous, which makes things more difficult. For now, we have just discretized the space of states. Later on, we'll try to use the local approximation technique [1]. This method consists in computing the benefit for some states on a grid. If we meet states that are not on the grid, we compute its benefit using distances functions to the grid.

B. The action space

At each time step, we choose one action made of two decisions :

- $a^{(1)}$ the quantity of electricity supplied (or taken from) the battery,
- $a^{(2)}$ the quantity of electricity supplied to the electrolyzer.

The other energy transfers are fixed by the two following equations :

$$E_{\text{prod}} = E_{\text{sold}} + E_{\text{to battery}} + E_{\text{to electrolyzer}} \tag{1}$$

$$Q_{\text{prod}} = Q_{\text{to vessel}} + Q_{\text{to clients}} \tag{2}$$

$E_{\text{to clients}}$ is fixed. E_{prod} is independent of our will, and is determined by the weather. $E_{\text{to battery}}$ and $E_{\text{to electrolyzer}}$ are determined by our policy. Fixing the input power to the electrolyzer is equivalent to fixing the ouput mass flow of H_2 : Q_{prod}. Finally, $E_{\text{sold to grid}}$ is fixed by equation (1), and $Q_{\text{to vessel}}$ is fixed by equations (2)

Same as for the state space, we have discretized our action space. It will also be constrained by the physics of our problem : the charge of the battery and the electrolyzer production rate should not change too much between two timesteps. This will limit the number of available actions at each time step.

C. Defining the transitions

Now, let's define the transition function $T(s'|s, a)$, which assigns the probability of reaching state s' by choosing action a at state s.

In our case, this transition is not very complex because :

- $s^{(1)}$ is deterministically incremented : $s^{(1)\prime} = s^{(1)} + \tau$, where τ is the timestep of the simulation,
- $s^{(4)}$ and $s^{(5)}$ are deterministically computed using the latest action: $s^{(4)\prime} = s^{(4)} + a^{(1)}$ and $s^{(5)\prime} = \eta * a^{(2)}/\tau$, where η is the efficiency of the electrolyzer [kg/kWh]
- $s^{(6)}$ is fixed by the previous quantity in the vessel plus what we have produced meanwhile, meanwhile what we provide to our clients: $s^{(6)\prime} = s^{(6)} + s^{(5)} * \tau - Q_{\text{to clients}}$
- $s^{(2)}$ and $s^{(3)}$ change stochastically, but do not depend on the action. As a simplification, we assume they only depend on the time of the day $s^{(1)}$. These transitions probabilities are approximated using Monte Carlo estimates on our data

D. Defining the reward

The reward function should enable our agent to learn the task. Thus :

- It must encourage profitable behaviors
- It must penalize undesirable states (e.g an electrolyzer production rate below 20% or battery fully discharged).

There is no need for the reward to be very complex for now. By just rewarding the agent proportionally to the profit it makes by selling the electricity, while penalizing dangerous or non physical states (battery state of charge ¡= 0), the agent should be able to learn on its own when to store, and when to sell or buy its electricity to maximize the total profit.

VI. Alternative approach: MILP optimization

Alternatively to using a MDP description, we could try to model our hydrogen supply chain as an optimization problem. The comparison of the two models will allow us to show the advantages of each approach and determine which one is best suited to optimize the management of the hydrogen supply chain for the company. We can already identify one weakness of the optimization, which can not include the real-time electricity market without implementing the problem using dynamic programming to update the price of the electricity on this market on an hourly basis as it should be. Let's now present the mathematical formulation of the optimization approach to this problem.

A. Graph

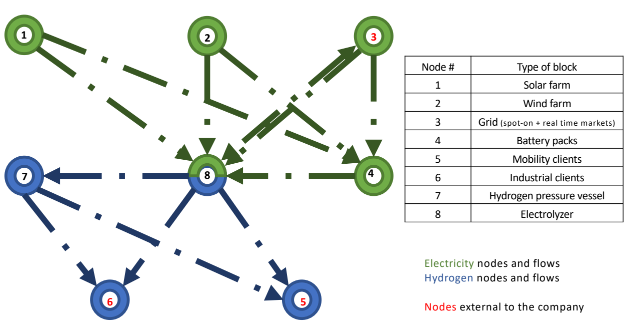

Fig. 2. Nodes and edges of the optimization problem

B. Input data

- $\Phi^{(i)}_{wind}$: Predicted power from the wind farm $[kW]$
- $\Phi_{solar}(i)$: Predicted power from the solar farm $[kW]$
- $\Psi^{(i)}_{mob}$:Hydrogen contracts for mobility $[kg(H_2)]$
- $\Psi^{(i)}_{ind}$: Hydrogen contracts for the industry $[kg(H^2)/hr]$
- $C_{day-ahead}$: Predicted electricity prices on the day-ahead market $[\$/kWh]$

C. Defining the variables

1) Decision variables:

- $\Phi_{batt}(i)$: Net power feeding the battery packs $[kW]$
- $\Psi_{vess}(i)$: Net hydrogen flow feeding the hydrogen vessel $[kg(H_2)/hr]$
- $\Psi^{(i)}_{prod}$: Production rate of the electrolyzers $[kg(H_2)/hr]$
- $\Phi_{grid}(i)$: Net power bought from the grid $[kW]$

2) Other variables:

- $\Phi^{(i)}_{(j1,j2)}$: Power from node $j1$ to node $j2$
- $\epsilon^{(i}_{batt}$: Load level of the battery packs $[kWh]$
- $m^{(i)}_{vess}$: Load level of the hydrogen vessels $[kg(H_2)]$
- $y^{(i)}_{prod}$: Electrolyzer On/Off (binary) $[]$
- $R^{(i)}$: Net revenue from electricity sales $[]$
- $\Phi^{(i)}_{ind)}$: Power sold to the industry $[kWh/hr]$
- $\Psi^{(i)}_{prod}$: Hydrogen production rate $[kg(H_2)/hr]$

D. Parameters of the model

1) Model of the simulation:

- Δt : duration of one iteration $[hr]$ (1hr in our case)
- N_t : number of iterations

2) parameters of the electrolyzers:

- η_{conv} : optimal efficiency of conversion (at max production rate) $[kWh/kg(H_2)]$
- Φ^{max}_{prod} : max hydrogen production capacity $[kWh/kg(H_2)]$
- Φ^{min}_{prod} : min hydrogen production capacity $[kWh/kg(H_2)]$ (the electrolyzer is shut-down otherwise)

3) parameters of the hydrogen pressure vessel:

- Ψ^{max}_{vess} : Maximum charge/discharge rate of the pressure vessel (constrained by H_2 compressors) $[kg(H_2)/hr]$
- q^{nom}_{vess} : Nominal storage capacity $[kg(H_2)]$
- C^{max}_{vess} : Coefficient of max vessel load
- C^{min}_{vess} : Coefficient of min vessel load

4) Parameters of the battery packs:

- ϵ^{nom}_{batt} : Nominal energy storage capacity $[kWh]$
- C^{max}_{batt} : Coefficient of max battery load
- C^{min}_{batt} : Coefficient of in battery load
- Φ^{max}_{batt} : Max charge/discharge power $[kW]$
- $C_{batt-penalty}$: Penalty coefficient associated with the dynamic usage of the electrolyzer $[\$/(kW/hr)]$
- η_{batt} : efficiency of storage (round-cycle)

E. Constraints

1) Power balance at each node:

- Wind farm (node 1)
  $\Phi^{(i)}_{wind} = \Phi^{(i)}_{(1,4)} + \Phi^{(i)}_{(1,8)}$
- Solar farm (node 2)
  $\Phi^{(i)}_{solar} = \Phi_{(2,4)} + \Phi^{(i)}_{(2,8)}$
- The grid (node 3)
  $\Phi^{(i)}_{grid} = \Phi_{(3,4)} + \Phi^{(i)}_{(3,8)}$
- Battery packs (node 4)
  $\Phi^{(i)}_{batt,int} = \Phi^{(i)}_{(1,4)} + \Phi^{(i)}_{(2,4} + \Phi^{(i)}_{(3,4)}$
  $\Phi^{(i)}_{batt,out} = \Phi^{(i)}_{(4,6)} + \Phi^{(i)}_{(4,8)}$
- Electrolyzer (node 8)
  $\Phi^{(i)}_{prod} = \Phi_{(1,8)} + \Phi^{(i)}_{(2,8)} + \Phi^{(i)}_{(3,8)} + \Phi^{(i)}_{(4,8)}$

2) Hydrogen balance at each node:

- Contracts for mobility (hydrogen stations) (node 5)
  $\Psi^{(i)}_{mob} = \Psi^{(i)}_{(7,5)} + \Psi^{(i)}_{(8,5)}$
- Contracts with Industrial clients (node 6)
  $\Psi^{(i)}_{ind} = \Psi^{(i)}_{(7,6)} + \Psi^{(i)}_{(8,6)}$
- Hydrogen pressure vessel (node 7)
  $\Psi^{(i)}_{vess} = \Psi_{(7,5)} + \Psi^{(i)}_{(7,6)}$
- Electrolyzer (node 8)
  $\Psi^{(i)}_{prod} = \Psi_{(8,5)} + \Psi_{(8,5)} + \Psi_{(8,7)}$

3) Update load of storage units:

- Battery pack (node 4)
  $\epsilon^{(i+1)}_{batt} = \epsilon^{(i)}_{batt} + (\Phi^{(i)}_{(1,4)} + \Phi^{(i)}_{(2,4)} + \Phi^{(i)}_{(3,4)} - \Phi^{(i)}_{(4,8)})\Delta t$
- Hydrogen pressure vessel (node 7)
  $m^{(i+1)}_{vess} = m^{(i)}_{vess} + (\Psi^{(i)}_{(8,7)} - \Psi^{(i)}_{(7,5)} - \Psi^{(i)}_{(7,6)})\Delta t$

4) Constraints on the H_2 production:

- Simplified model - we here assume that the efficiency of the conversion occurring in the electrolyzer is constant (while it actually depend on $\Phi_{prod}^{(i)}$):
$\Phi_{prod}^{(i)} = 1/(\eta_{conv}) * \Psi_{prod}^{(i)}$
- Improved model - we keep a form of dependence of the efficiency of conversion on the load of the electrolyzer:
(1) 1st case: if $\Phi_{prod}^{(i)} > 3/4 * \Phi_{prod}^{nom}$ then the efficiency is at it's nominal value
$\Phi_{prod}^{(i)} * y_1^{(i)} > 3/4 * \Phi$
$\Psi_{prod}^{(i)} \leq 1/(\eta_{conv}) * \Phi_{prod}^{(i)}$
(2) 2nd case: if $\Phi_{prod}^{(i)} < 3/4 * \Phi_{prod}^{nom}$ then the efficiency is reduced to 85% of it's nominal value
$\Phi_{prod}^{(i)} * y_2^{(i)} > 1/2 * \Phi$
$\Psi_{prod}^{(i)} \leq 1/(0.85 * \eta_{conv}) * \Phi_{prod}^{(i)} + y_1^{(i)} * 2/(\eta_{conv}) * \Phi_{prod}^{norm}$
(3) 3rd case: if $\Phi_{prod}^{(i)} < 1/2 * \Phi_{prod}^{nom}$ then the efficiency is reduced to 65% of it's nominal value
$\Phi_{prod}^{(i)} * y_3^{(i)} > 1/2 * \Phi_{prod}^{norm}$
$\Psi_{prod}^{(i)} \leq 1/(0.65 * \eta_{conv}) * \Phi_{prod}^{(i)} + y_2^{(i)} * 2/(\eta_{conv}) * \Phi_{prod}^{norm}$
(4) 4th case: if $\Phi_{prod}^{(i)} < 1/4 * \Phi_{prod}^{nom}$ then the efficiency is reduced to 50% of it's nominal value
$\Psi_{prod}^{(i)} \leq 1/(0.5 * \eta_{conv}) * \Phi_{prod}^{(i)} + y_3^{(i)} * 2/(\eta_{conv}) * \Phi_{prod}^{norm}/(0.5 * \eta_{conv}) * \Phi_{prod}^{(i)}$

F. Objective function

The objective function corresponds to the revenue earned by the company, which is made of the net revenue from selling electricity to the industry and the grid as well as the revenue from selling hydrogen. However the sales of hydrogen to the mobility and the industry are fixed in advance, so the revenue earned from selling hydrogen is fixed and thus doesn't need to be included in the objective function.
$J = \sum_{i=1}^{N_t} (\Phi_{(8,6)}^{(i)} - \Phi_{(3,8)}^{(i)} - \Phi_{(3,4)}^{(i)}) * C_{day-ahead}^{(i)}$

VII. Contributions

So far, the work has been divided as such : Zoe found the data and wrote the code to extract them. Calvin defined the physical models for the different subsystems of our environment (e.g electrolyzer, batteries etc..), as well as wrote the code for the MILP optimization part. Finally, Malik coded the environment for the reinforcement learning simulation. The code is available at https://github.com/mboudiaf/CS229-project

References

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

[2] Yusheng Luo, Min Xian, Manish Mohanpurkar, Bishnu P. Bhattarai, Anudeep Medam, Rahul Kadavil and Rob Hovsapian
Optimal Scheduling of Electrolyzer in Power Market with Dynamic Prices , 2018

[3] P. Kofinasa,b,, A.I. Dounisb, G.A. Vourosa *Fuzzy Q learning for multi agent decentralized energy management in microgrids*

[4] Elia, Belgiums electricity transmission system operator
http://www.elia.be/en/grid-data/power-generation

[5] SMARD, Strommarktdaten, electricity market data
https://www.smard.de/en/