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 :

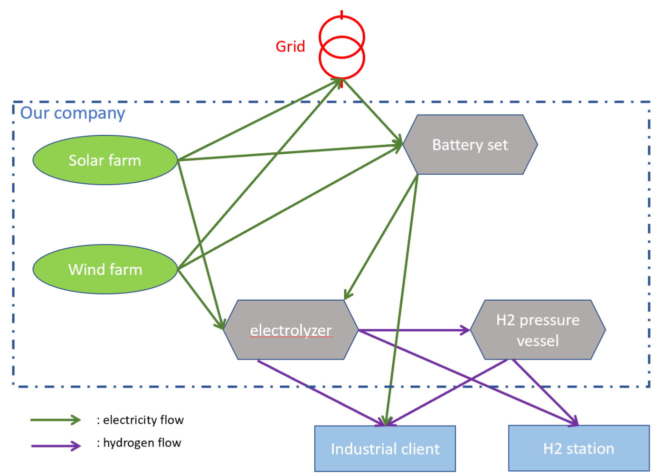

Fig. 1. Vanilla hydrogen/electricity supply chain

- *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 ? Actually, we are not even able to say whether a particular action taken at a particular time is judicious regarding the whole simulation. Instead, we're able to judge higher features of an action like : a sequence of actions is good if it yields a lot of money (in comparison to other baseline policies). Hence, it sounds intuitive to work in the reinforcement learning framework.

Thus, we will try three different types 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.

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*