Learning to Run a Power Network Energies of the future and carbon neutrality

Gaëtan Serré
gaetan.serre@universite-paris-saclay.fr

1 Algorithm & Results

The goal of this exercise was to provide a reinforcement learning solution to the challenge *L2RPN NEURIPS 2020*[1]. I chose to implement a Deep Q-Learning algorithm. For reasons of time and computational power, I was able to train my model during 1,000 episodes considering only 50 actions (including the *do-nothing* action). I used the default reward. We can see on 1 that the more the number of episodes increases, the longer the model survives

Episode	Average steps survived	Average total reward
0-100	16	6,000
100-300	32	20,000
400-500	100	100,000
600-700	500	500,000
800-1000	900	900,000

Table 1: Summary of the results of the training phase

My model reaches a score of -2.432371 on the competition, which is worse than the *do-nothing* strategy, however, with more episodes and some optimizations on the neural network used or the selection of the power grid features used to train it, I think my algorithm might have a better score.

2 Difficulties

During these few days, I encountered some difficulties.

- The first one was first of all understand the purpose of the challenge and what was the stake of the problem because I was not familiar with power grid. The starting kit and the paper *Learning to run a Power Network Challenge: a Retrospective Analysis* [2] helped me a lot.
- Another difficulty was to use the *grid2op* library. I had to do a lot of testing to find out what was available to me to make a reinforcement learning model. The starting kit also helped me a lot.
- The choice of the model was quite natural, I had already implemented a Q-Learning algorithm and I thought that Deep Q-Learning was very suitable for the problem. However, I had never implemented it before so I had to follow some tutorials.
- Features selection was also a problem. Beyond a version error of *grid2op* that gave me a hard time, looking at the public submissions, I could see that the best teams were selecting features of the power grid. As I don't have the expert knowledge to do the same, I decided to keep them all but it is surely not the most optimized solution.
- The space action having more than 70,000 possible actions, I had to select only a subset of them in order to save computing time. I was able to use the file *top1000_ actions.npz* provided by the winning team [3]. This file contains the 1,000 best actions according to them. I chose to keep only the first 50 while being sure that the *do-nothing* action is included. This way, I knew that I could never have the best solution but I could do as good as the *do-nothing* agent.
- The last difficulty was to participate in a more "professional" competition than the ones I had participated in. More specifically, formatting my submissions took a lot of time.

3 What I have learned

By surpassing the difficulties mentioned above, I learned to implement a Deep Q-Learning algorithm and to participate in a professional competition. Thanks to *grid2op*, I was able to learn the vocabulary and interfaces of reinforcement learning. My knowledge in this domain came from personal projects in which I implemented everything *from scratch*, so I didn't have this general culture. Finally, I had an overview of the problem of managing power grids and what is at stake for our society.

4 Conclusion

Before I even started, I was already very motivated by this internship. I learned about reinforcement learning through its application in game theory and I have been very interested in it since then. Thanks to this challenge, I discovered a more "realistic" application than Go or chess of reinforcement learning and I really liked it. I would like to go deeper into the problem by improving my algorithm or or by trying others.

References

[1] L2rpn neurips 2020 - adaptability track. website.

[2] A. Marot, B. Donnot, G. Dulac-Arnold, A. Kelly, A. O'Sullivan, J. Viebahn, M. Awad, I. Guyon, P. Panciatici, and C. Romero. Learning to run a power network challenge: a retrospective analysis, 2021.

[3] B. Zhou, H. Zeng, Y. Liu, K. Li, F. Wang, and H. Tian. Action set based policy optimization for safe power grid management. 2021.