Deep Reinforcement Learning to Minimize Long-Term Carbon Emissions and Electricity Cost in the Investment of Electricity Generation

Anonymous Author(s)
Affiliation
Address
email

Abstract

A transition from a high-carbon emitting electricity power system to one based on renewables would aid in the mitigation of climate change. Decarbonization of the electricity grid would allow for low-carbon heating, cooling and transport. Investments in renewable energy must be made over a long time horizon, with uncertainties in future electricity demand and costs to consumers and investors.

To account for imperfect information of the future, we use the deep deterministic policy gradient (DDPG) deep reinforcement learning approach to optimize for a low-cost, low-carbon electricity supply using a modified version of the FTT:Power model. In this work, we model the UK and Ireland markets. The DDPG algorithm is able to learn the optimum electricity mix through experience and achieves this between the years of 2017 and 2050. We find that a transition from fossil fuels and nuclear power to renewables, based upon wind, solar and wave would provide a cheap and low-carbon alternative.

1 Introduction

A transition from a high carbon electricity supply to a low-carbon system is central to avoiding catastrophic climate change [1]. A low-carbon electricity supply will aid in the decarbonization of the automotive and heating sectors. Such a transition must be made in a gradual approach to maintain grid reliability [2].

Renewable energy costs, such as solar and wind energy, have reduced over the last ten years, making them cost-competitive with fossil fuels. These price drops are projected to continue [3]. The future cost of generation, demand and fuel prices, however, remain uncertain over the long-term future. These uncertainties are risks which investors must analyze while making long-term decisions.

In this paper, we use the deep deterministic policy gradient (DDPG) reinforcement learning algorithm to simulate the behaviour of investors over a 33-year horizon, between 2017 and 2050 [4]. The model is parameterized and begins in 2007, however, the investment decisions begin in 2017. We begin in these years due to the prior parameterization of the FTT:Power model with historical data up until this time. We projected until 2050 due to the fact that this is a frequent target for governments to reach zero carbon. The environment used was a modified version of the FTT:Power model [5]. FTT:Power is a global power systems model that uses logistic differential equations to simulate technology switching.

We modified the FTT:Power model to use the DDPG algorithm in place of the logistic differential equations to simulate investment decisions. In addition to this, we simulated two countries: the United Kingdom and Ireland. This was achieved through the use of the same FTT:Power model. We

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

choose these due to the wealth of prior work on these countries which we can use for comparison [6, 7]. The DDPG algorithm allows us to simulate the decisions made by investors under imperfect information, such as future electricity costs, taxes and demand. This work enabled us to see whether a low-carbon mix is possible over the next 33-years to avert climate change.

Prior work in this domain has tackled the capacity expansion problem. For example, Oliveira *et al.* also use reinforcement learning for the capacity expansion problem [8]. Whilst Oliveria *et al.* provide detailed calculations of agents for the capacity expansion problem, we reduce this complexity to a series of observations of the environment, to allow for an emergent behaviour.

Kazempour *et al.* use a mixed integer linear programming approach to solve the generation investment problem [9]. Our approach is in contrast to a mixed-integer linear programming problem, where full knowledge of the time-horizon is required. In our work, we address a gap in the literature for the capacity expansion problem over a 33 year time period using deep reinforcement learning to reduce both carbon emissions and electricity price.

Through this work, it is possible to assess whether a low cost, low-carbon electricity mix is viable over the long-term using a deep reinforcement learning investment algorithm, as well as finding what this optimum mix should be. This work enables us to closely match the investment behaviour of rational agents, without knowledge of the future. It can help guide investors on which technologies to invest in over the long-term, as well as the proportions to invest.

2 Model and methodology

In this paper, we used the Future Technology Transformations for the power sector model (FTT:Power) [5]. This model represents global power systems based on market competition, induced technological change and natural resource use and depletion. This technological change is dependent on previous cumulative investment [5]. The model uses a dynamic set of logistic differential equations for competition between technology options.

For this work, we modified the FTT:Power model to use the deep reinforcement learning investment algorithm, DDPG. That is, the size of the investment made in each technology was made by the DDPG algorithm. In addition, we reduced the model to only consider the countries of Ireland and the UK. This enabled us to iterate through enough episodes for the reinforcement learning to converge to an optimal reward. With more time, however, it would be possible to undertake this optimisation for the whole world.

Reinforcement Learning

The investment decision-making process can be formulated as a Markov Decision Process (MDP) [10]. In an MDP environment, an agent receives an observation about the state of their environment s_t, chooses an action a_t and receives a reward r_t as a consequence of their action and the resultant change on the environment. Solving an MDP consists of maximizing the cumulative reward over the lifetime of the agent.

For our simulation environment, the agent makes a continuous investment decision for each energy technology, in each region and each year, starting from 2017 until 2050. Technology switching is modelled using a pairwise comparison of flows of market shares of different electricity generation capacity. That is, how much capacity flows from one technology to another.

The agent's observation space is a vector consisting of the electricity produced by each technology, total capacity, total CO_2 emissions over the simulation, levelized cost of electricity (LCOE) including both taxes and without taxes, cumulative investment in each technology, investment in new capacity, carrier prices by commodity, fuel costs and carbon costs.

The reward r is defined as:

$$r = -\left(1000 \times \mathrm{CO}_{2_e} + \frac{LCOE}{1000}\right), \tag{1}$$

where CO_{2_e} is equal to total CO_2 emissions over the simulation The LCOE is calculated without taxes. The scaling factors are used to place the $LCOE$ and CO_2 on the same scale. The reward was

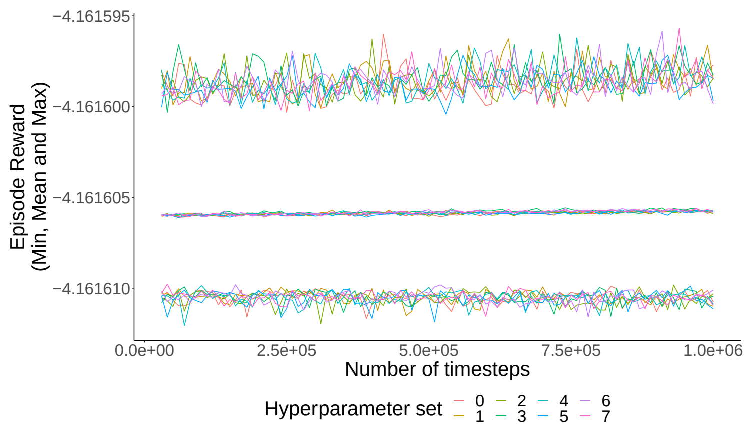

Figure 1: Training with different hyperparameters, displaying the minimum, mean and maximum rewards per episode.

multiplied by -1 due to the RL algorithm maximizing reward and our requirement to reduce both LCOE and CO_2 emissions.

RL approaches have been used to solve MDP through a trial and error based approach [11]. In recent times RL has been extended to incorporate Deep Reinforcement Learning (DRL). DRL relies on deep neural networks to overcome the problems of memory and computational complexity [12].

We applied the deep deterministic policy gradient (DDPG) DRL algorithm [4] from the Ray RLlib package [13]. The DDPG algorithm is made up of an actor and critic network. We designed both of these to have two hidden layers, made up of 400 and 300 units per layer. The training batch size was set to 40,000. We chose these parameters as they were the default implementation in Ray RLlib. We trialled a variety of different configurations for number of neurons per layer. To increase speed of computation, we reduced the simulation to run from 2007 to 2017. We chose this number as it would allow for a transition in electricity mix. However, we found that the approach worked well. We trialled the use of two and three layers, as well as varying permutations of 300, 400 and 500 neurons.

3 Results

Our results show that our investment agent is able to increase its reward over time, as shown in Figure 2. A total of ∼400,000 steps were required to see a levelling off in reward. The total time taken to simulate ∼400,000 steps was ∼8 days. We stopped the training and simulation after this time due to diminishing returns and the cost of computation.

Figure 3 displays the results of reinforcement learning algorithm. Before the black vertical line (2017), the investments made are based upon historical data used by FTT:Power. The reinforcement learning algorithm starts to make investments after the black vertical line.

The historical electricity mix before 2017 is largely based on fossil fuels: coal, combined cycle gas turbine (CCGT) and oil. Additionally, nuclear is a major component of the electricity mix before 2009. After reinforcement learning optimizes for LCOE and carbon emissions, a rapid transition occurs from fossil fuel and nuclear to renewable energy.

This rapid transition occurs due to the reinforcement learning algorithm not taking into account the technical and timeframe constraints embedded in the unmodified FTT:Power model. However, although it is likely that whilst the transition speed is unrealistic, the electricity mix found by the reinforcement learning algorithm is likely to be optimal, according to the reward function defined in Equation 1.

The primary source of energy after the reinforcement learning algorithm begins is offshore, followed by onshore, solar photovoltaics (PV) and wave. As can be seen by Figure 4, the carbon emissions reduce significantly at the time that the reinforcement learning algorithm begins to control investments.

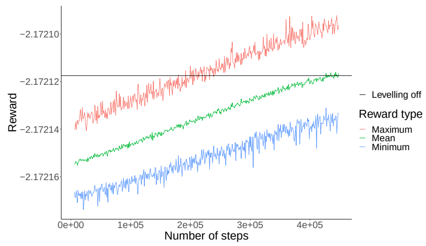

Figure 2: Mean, minimum and maximum rewards over run time.

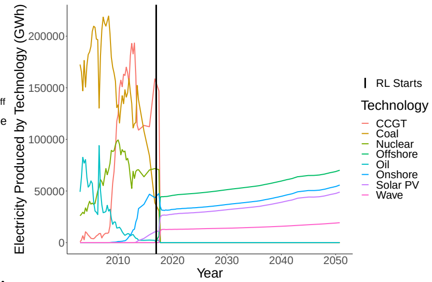

Figure 3: Electricity mix over time.

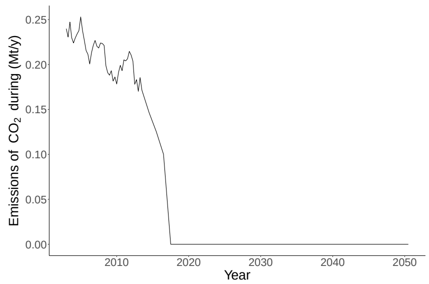

Figure 4: Carbon emissions.

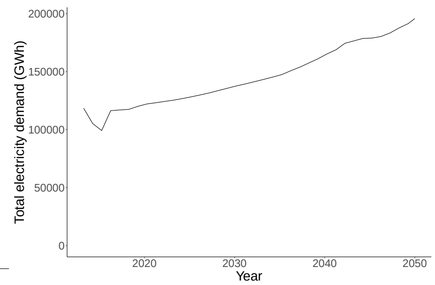

Figure 5: Demand scenario over simulation.

This mix of renewable electricity generation across Ireland and the UK allows for demand to be met during the quarterly time periods of the model. The demand scenario is shown in Figure 5, where the demand can be seen to closely match the electricity mix shown by Figure 3.

4 Discussion

A transition from a high carbon-emitting electricity grid to a low-carbon system is required. In order to achieve this, investments in electricity generators must be made whilst taking into account future uncertainty. In this paper, we have modelled a central agent which makes investment decisions in an uncertain environment to find an optimal low-cost, low-carbon electricity mix. To achieve this, we used the reinforcement learning algorithm, DDPG. The environment is modelled using FTT:Power.

Through this exercise, we are able to see the optimal electricity mix in the UK and Ireland. We found that a mixture of renewable sources such as wind, solar and wave power would meet demand at quarter year intervals, as well as providing a cost-effective and low-carbon system.

A limitation of this work is the fact that the investment algorithm does not take into account the technical and timeframe constraints of transitions between technologies. It is for this reason that the reinforcement learning algorithm is able to make such a rapid transition in 2017. However, we believe that the investment algorithm is able to find a general solution to the problem of investing in a cost-efficient and low-carbon system over a long time horizon.

In future work, we would like to increase the number of steps of the FTT:Power model to more adequately model the investment behaviour introduced by the reinforcement learning algorithm. A lower number of simulated time steps leads to an overestimation of the supply of renewables and underestimation of storage and dispatchable technologies [14]. In addition, an increase in the number of countries modelled would enable us to see a global picture of how different, interdependent regions may evolve in a new climate of a requirement of low-carbon emissions. This would require an exponentially longer runtime for the reinforcement learning algorithm to converge. This is due to the increased number of decisions that the reinforcement learning algorithm would need to make to account for the different countries. Finally, we would like to incorporate the technical and timeframe constraints for technology switching. This could be undertaken by modifying the reward function to ensure the transition remains within these constraints.

5 Acknowledgements

Acknowledgements redacted to maintain anonymity for double-blind.

References

[1] A. J. M. Kell, M. Forshaw, and A. S. McGough, "Long-Term Electricity Market Agent Based Model Validation using Genetic Algorithm based Optimization," *The Eleventh ACM International Conference on Future Energy Systems (e-Energy'20)*, 2020.

[2] F. Kahrl, J. Williams, D. Jianhua, and H. Junfeng, "Challenges to China's transition to a low carbon electricity system," *Energy Policy*, vol. 39, no. 7, pp. 4032–4041, 2011.

[3] IEA, "Projected Costs of Generating Electricity," p. 215, 2015.

[4] J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous learning control with deep reinforcement," *ICLR*, 2016.

[5] J. F. Mercure, "FTT:Power A global model of the power sector with induced technological change and natural resource depletion," *Energy Policy*, vol. 48, pp. 799–811, 2012.

[6] L. M. H. Hall and A. R. Buckley, "A review of energy systems models in the UK : Prevalent usage and categorisation," *Applied Energy*, vol. 169, pp. 607–628, 2016.

[7] N. Hughes and N. Strachan, "Methodological review of UK and international low carbon scenarios," *Energy Policy*, vol. 38, no. 10, pp. 6056–6065, 2010.

[8] F. S. Oliveira and M. L. Costa, "Capacity expansion under uncertainty in an oligopoly using indirect reinforcement-learning," *EJOR*, vol. 267, no. 3, pp. 1039–1050, 2018.

[9] S. J. Kazempour, A. J. Conejo, and C. Ruiz, "Strategic generation investment using a complementarity approach," *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 940–948, 2011.

[10] M. L. Puterman, "Markov decision processes: discrete stochastic dynamic programming," 2014.

[11] R. S. Sutton and A. G. Barto, "An introduction to reinforcement learning," *The MIT Press*, 2015.

[12] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A Brief Survey of Deep Reinforcement Learning," *IEEE Signal Processing Magazine*, pp. 1–16, 2017.

[13] E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J. E. Gonzalez, M. I. Jordan, and I. Stoica, "RLlib : Abstractions for Distributed Reinforcement Learning," 2014.

[14] S. Ludig, M. Haller, E. Schmid, and N. Bauer, "Fluctuating renewables in a long-term climate change mitigation strategy," *Energy*, vol. 36, no. 11, pp. 6674–6685, 2011.