Deep Reinforcement Learning to Minimize Long-Term Carbon Emissions and 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 the automotive sector with electric cars, amongst others. However, investments in renewable energy must be made over a long time horizon, with various uncertainties in future electricity demand and electricity costs.

To account for imperfect information of the future, we use the deep deterministic policy gradient (DDPG) deep reinforcement learning to optimize for a low-cost, low-carbon electricity supply using a modified version of the FTT:Power model. The DDPG algorithm is able to learn the optimum electricity mix through experience to achieve this goal 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]. Much of the work in decarbonization relies on a low-carbon electricity supply, such as cooling, heating and automotive, amongst others. Such a transition must be made in a gradual approach to avoid the frequent collapse of the electricity supply.

Renewable energy costs, such as solar and wind energy, have dropped recently, making them cost-competitive with fossil fuels. These drops in prices are projected to continue [2]. 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 investment 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 [3]. The environment used was a modified version of the FTT:Power model [4]. 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 only simulated two countries: the United Kingdom and Ireland. The DDPG algorithm allowed us to simulate the decisions made by investors under imperfect information over a 33-year period.

The reinforcement learning algorithm enabled us to model the behaviour of an investor without perfect knowledge of the future, with a view to reducing carbon emissions and the overall cost of the system. The reinforcement learning agent is a single actor that invests in both the UK and Ireland.

This work enabled us to see whether a low-carbon mix is possible over the next 33-years to avert climate change.

Oliveira *et al.* also use reinforcement learning for the capacity expansion problem [5]. They, however, focus on a 20-year time horizon. Kazempour *et al.* use a mixed integer linear programming approach to solve the generation investment problem [6]. In this 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. Our approach is in contrast to a mixed-integer linear programming problem, where full knowledge of the time-horizon is required. This work enables us to closely match the investment behaviour of rational agents, without a 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 in.

2 Model and methodology

In this paper, we used the Future Technology Transformations for the power sector model (FTT:Power). This model represents global power systems based on market competition, induced technological change and natural resource use and depletion. Induced technological change occurs through technological learning produced by cumulative investment and leads to nonlinear path dependent technological transitions [4]. The model uses a dynamic set of logistic differential equations for competition between technology options.

For this work, however, 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.

Reinforcement Learning

The investment decision-making process can be formulated as a Markov Decision Process (MDP) [7]. 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 based upon their environment and action. 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 matrix consisting of the electricity produced by each technology, total capacity, total CO_2 emissions over the simulation, levelized cost of electricity (LCOE) both including taxes and not including 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, and $LCOE$ is equal to LCOE, excluding taxes. The scaling factors were used to place the $LCOE$ and CO_2 on the same scale. The reward was 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 [8]. In recent times, however, RL has been extended to incorporate Deep Reinforcement Learning (DRL). DRL relies on deep neural networks to overcome the problems of memory complexity and computation complexity [9].

We applied the deep deterministic policy gradient (DDPG) DRL algorithm [3] from the Ray RLlib package [10]. 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 numbers due to their implementation in Ray RLlib and our inability to tune the hyper-parameters, due to an eight-day running time.

3 Results

Our results show that our investment agent is able to increase its reward over time, as shown in Figure 1. 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 2 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.

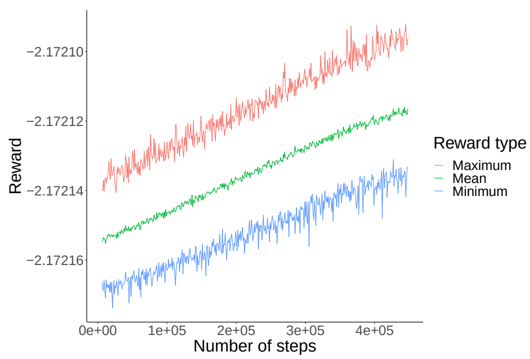

Figure 1: Mean, minimum and maximum rewards over run time.

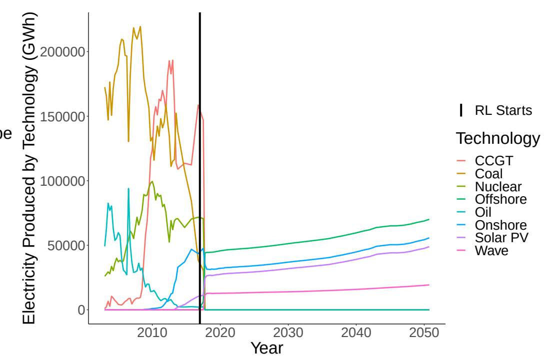

Figure 2: Electricity mix over time.

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 3, the carbon emissions reduce significantly at the time that the reinforcement learning algorithm begins to control investments.

This mix of renewable electricity generation across Ireland and the UK allows for demand to be met during the quarter year time periods of the model. The demand scenario is shown in Figure 4, where the demand can be seen to closely match the electricity mix shown by Figure 2.

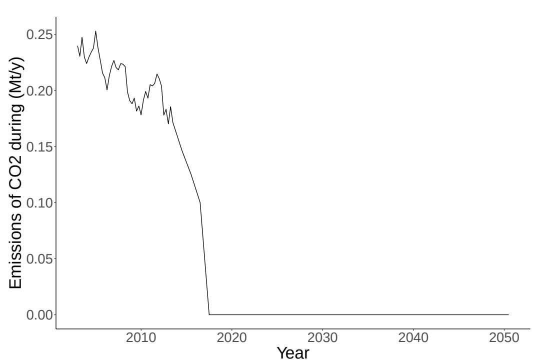

Figure 3: Carbon emissions.

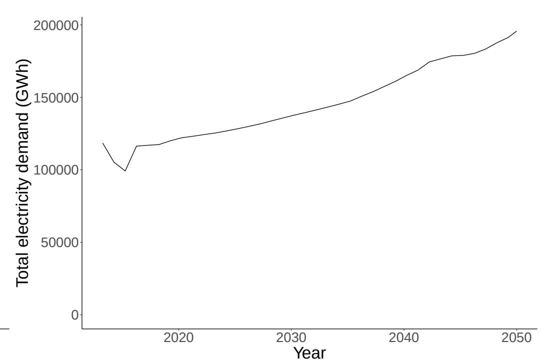

Figure 4: Demand scenario over simulation.

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 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 money owned and owed by the central decision-making agent. 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. 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.

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

[3] J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous learning control with deep reinforcement," *International Conference on Learning Representations (ICLR)*, 2016.

[4] 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.

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

[6] 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.

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

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

[9] 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.

[10] 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.