Uncertainty-Aware Reinforcement Learning for Electricity Demand Response

M.Sc. Thesis presentation

Ludwig Bald

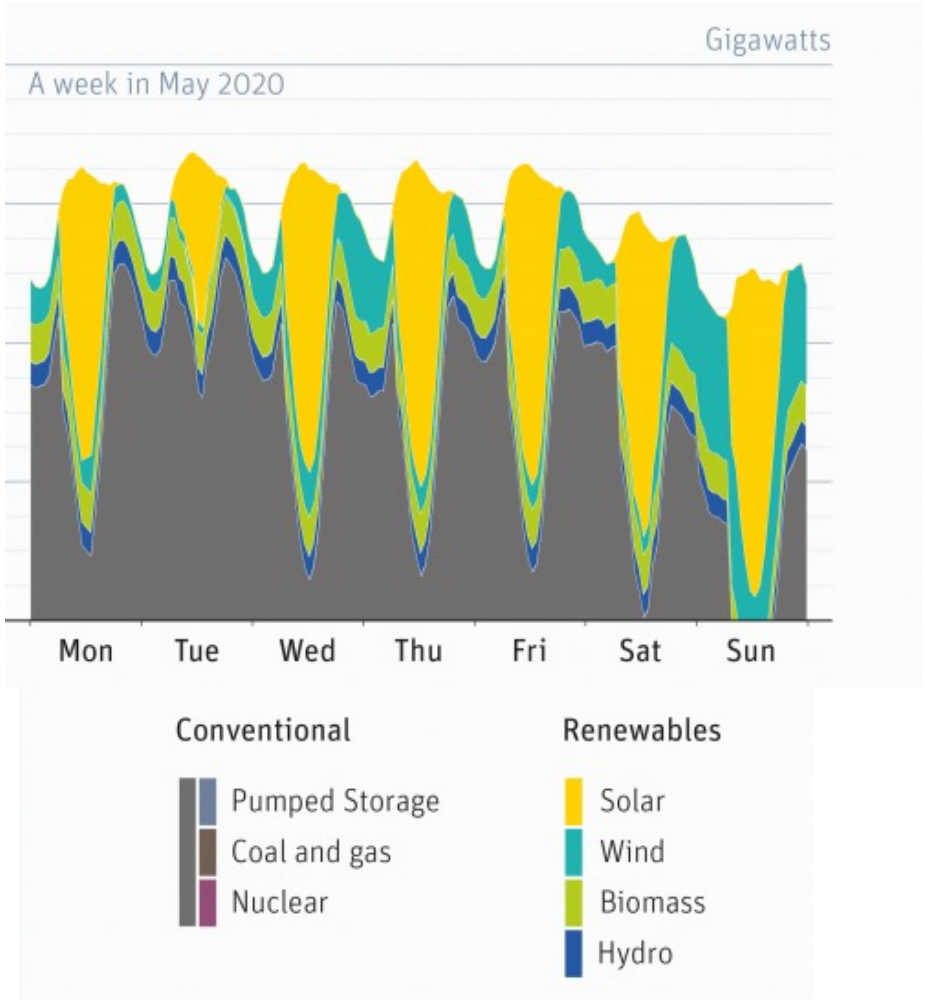
2022-12-19

Motivation – Methods – Experiments – Results - Discussion

Motivation: Flexibility

- Supply
- Dedicated Storage
  - expensive
- **Demand Response**
  - **Shift** demand through time
  - Reduce **peak loads**
  - How to coordinate?
  - Complex planning problem

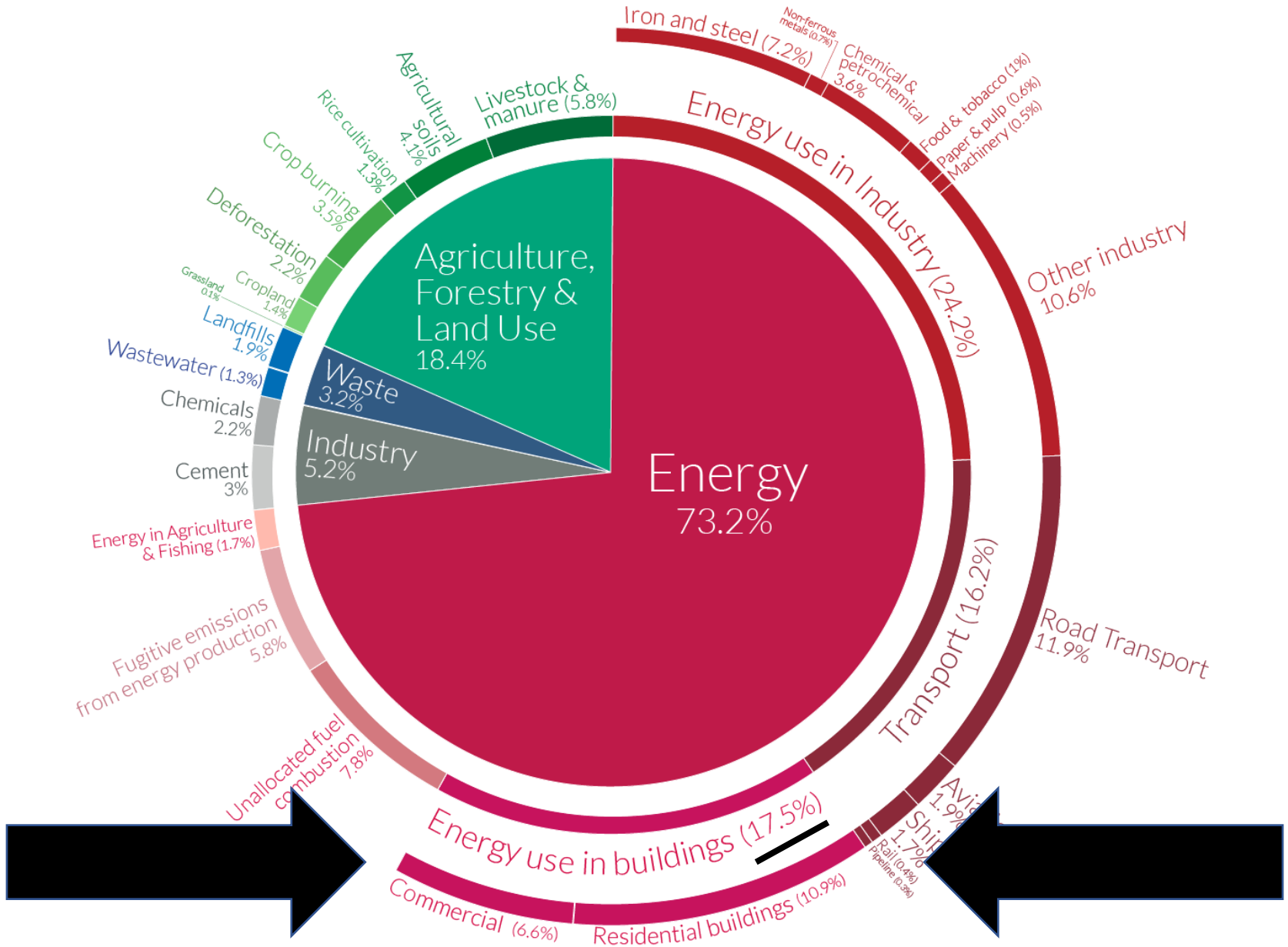
  → **Automation**



Global greenhouse gas emissions by sector

This is shown for the year 2016 – global greenhouse gas emissions were 49.4 billion tonnes CO_2eq.

Our World in Data

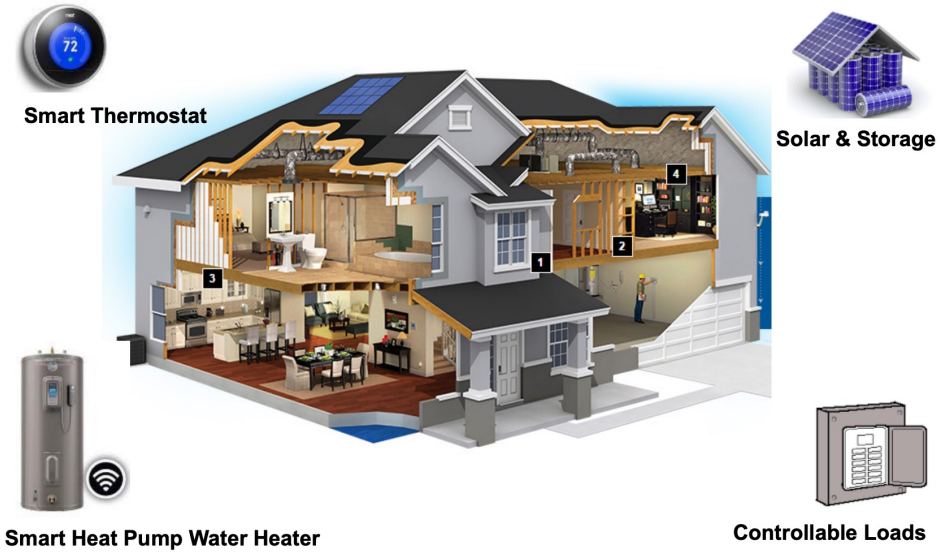


Methods

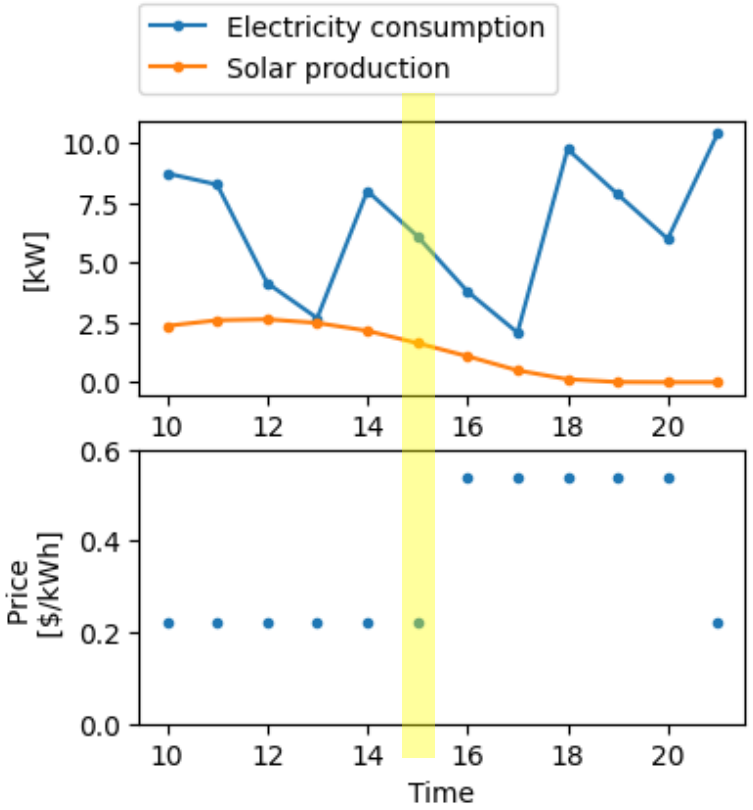
Methods: CityLearn

A 2019 Framework for RL Demand Response in **Buildings**

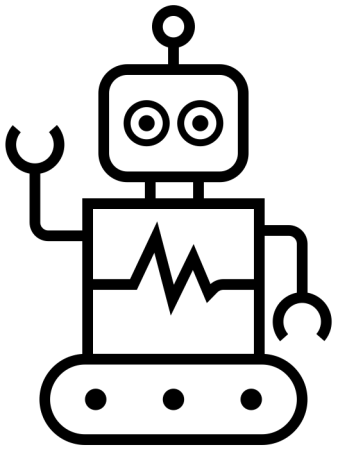
- Framework to control building-level **storage** with Reinforcement Learning
- Solar production and electricity consumption are real-world measured data
- 5 buildings in California, 1 year



CityLearn (Simplified)



Observation



Reward: New State:

Charge: -1$

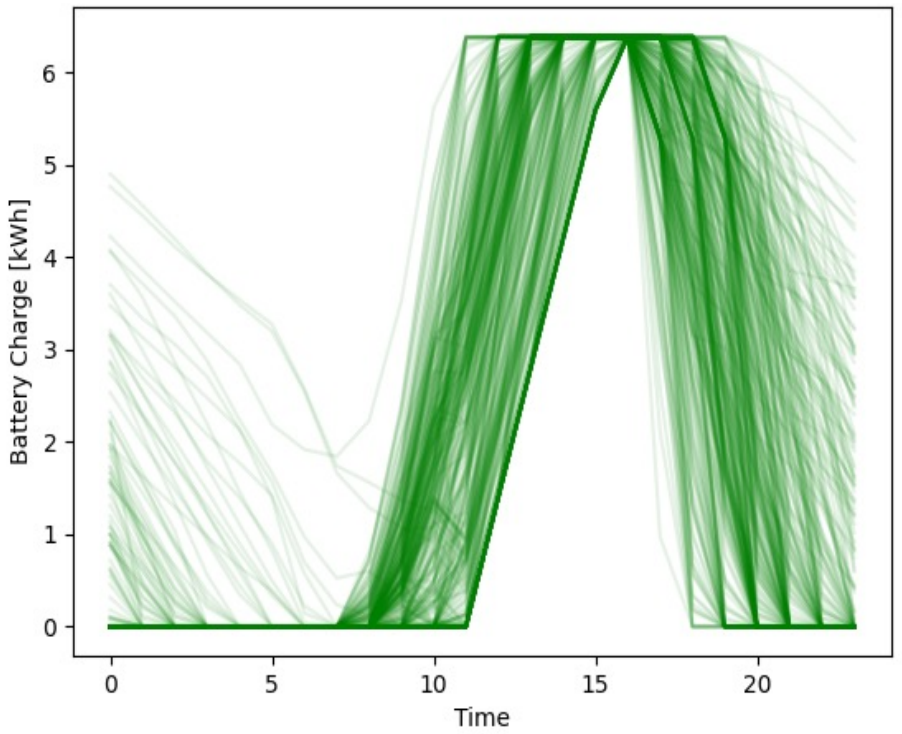
Future Savings possible

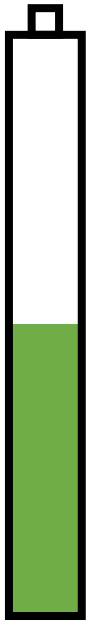
Release: -0$

No Future Savings

Handcrafted Rule-Based Agent

- Always store excess solar
- Always fill the battery before evening high prices





Metric:	Rule-Based	No Battery
Dollar Cost:	71.5%	100%
Carbon Emissions:	90.1%	100%
Average:	80.8%	100%

Reinforcement Learning (excerpt)

- Goal: Learn a **policy π^*** that recommends actions s.t. the **expected discounted reward** is maximized.
- In **Q-Learning**: Learn optimal state-action-values:

$$Q^*(s,a) = E_{\pi^*}\{r_t, \gamma r_{t+1}, \gamma^2 r_{t+2}, \dots | s_t = s, a_t = a, \pi = \pi^*\}$$

- Use Q-Values to select the optimal action

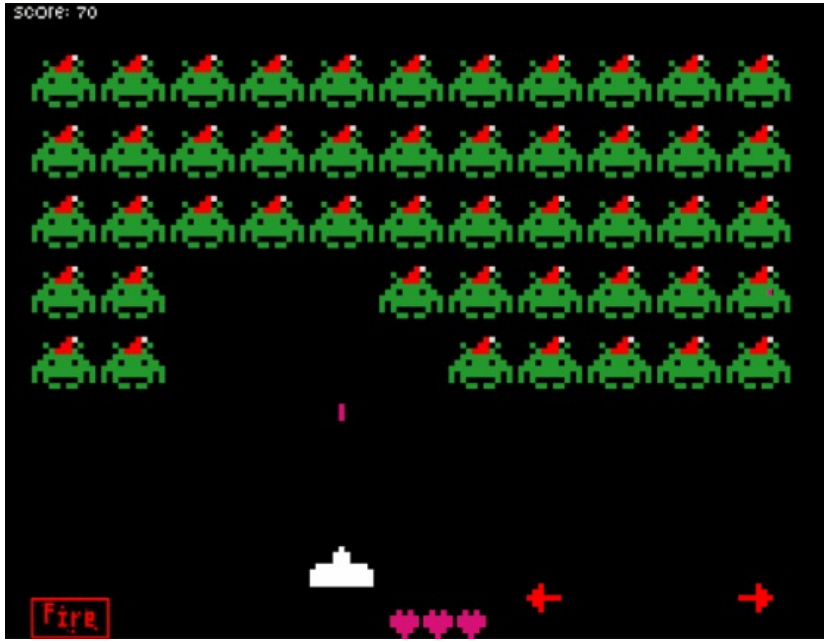
- Problem: Need to Generate Data

Q-Learning: Action Selection

- Q-value estimates can be improved by updating on any observation. (“Off-policy” learning)
- But: Some observations close to the optimal policy are **more important**.
- Idea: Use our approximate Q-values to select more important actions.
- ϵ-greedy

Deep Q-Network

- 2015: Able to play Atari Games!
- **Neural net**
  - Input: Observations
  - Output: **Q-value** per possible action.
- Further tricks:
  - ϵ-greedy Action Selection
  - Replay Buffer
  - TD-Learning



*Mnih et al., 2015

Uncertainty-Aware DQN

- Idea: Instead of modelling only the **expected** total reward, model a **distribution**!
- Two kinds of uncertainty of $\hat{R}$:
  - **Aleatoric uncertainty**: caused by known stochastic environment dynamics
  - **Epistemic uncertainty**: caused by not enough knowledge about environment dynamics
- Only Epistemic uncertainty is important

Experiments

Experimental Setup

- Test 3 algorithms:
  - DQN (**e-greedy** action selection)
  - DQN (**softmax** action selection)
  - **UA-DQN**
- **Data:** 1 Building, 1 year
- **Reward:** Normalized carbon emission + dollar cost
- **Tune** hyperparameters on 1 Building
- Record performance and action selection

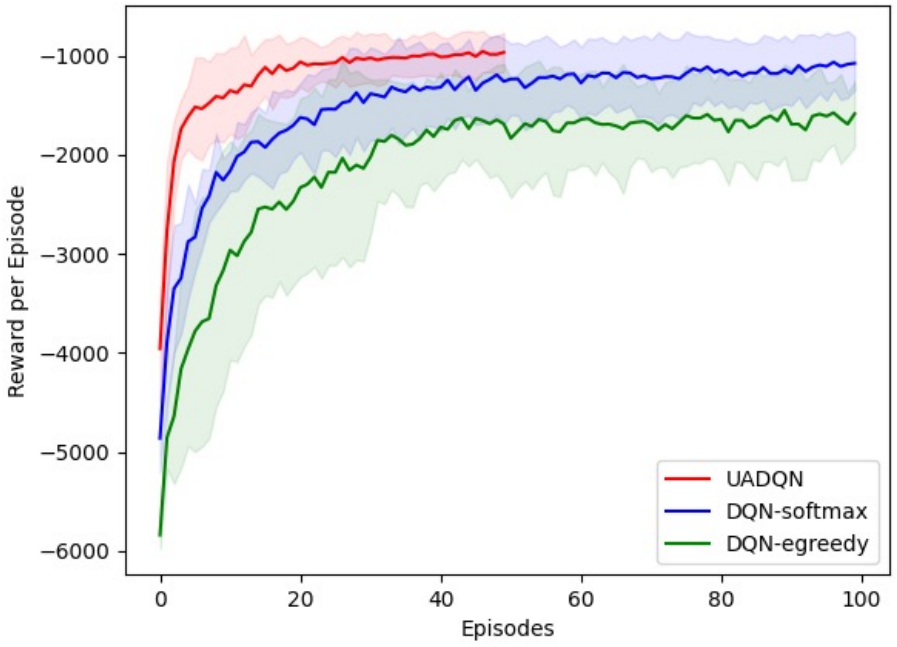
Hyperparameter Tuning

- Random Grid
- Tuned Parameters:
  - Learning rate
  - Batch size
  - Adam epsilon
  - E-Greedy epsilon+
  - Target network update frequency
- Repeat best runs for variance estimate
- Goal: Overfit on single building

Results

Results: Hyperparameter Tuning

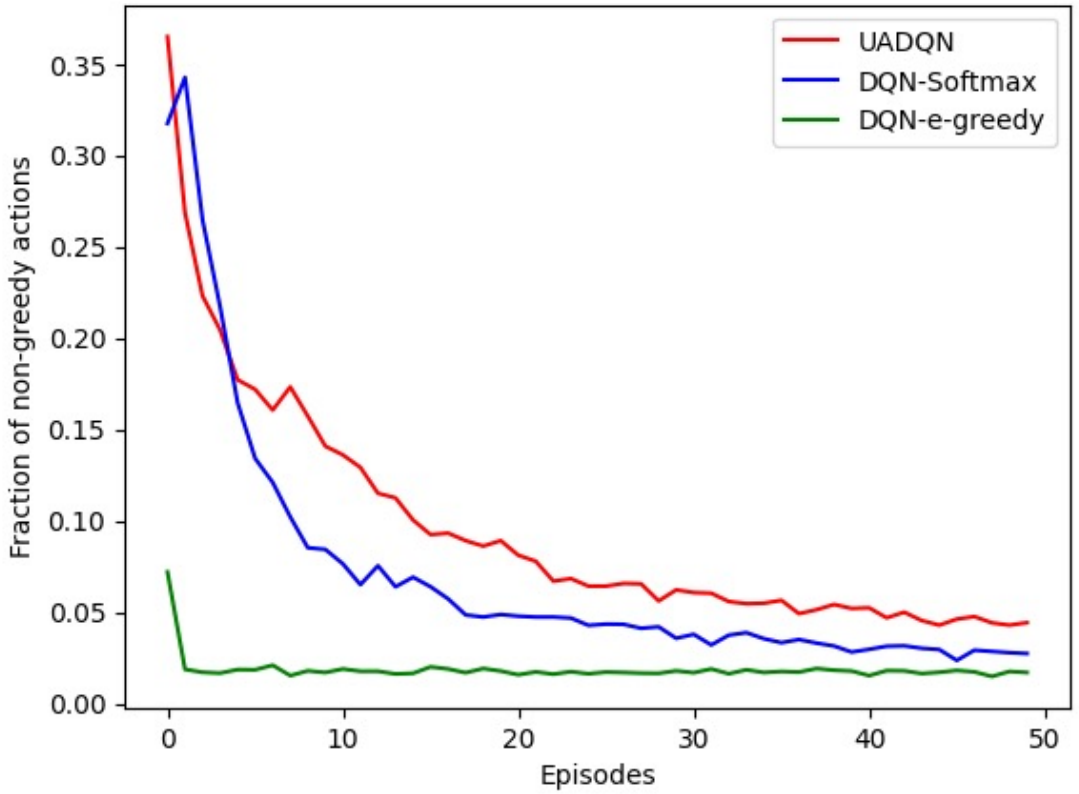
Reward



Greedy Performance

Agent	No battery	DQN-Softmax	DQN e-greedy	UA-DQN	Hand-crafted
Dollar Cost	1	0,83	0,82	0,82	0.72
Carbon Emission	1	0,93	0,93	0,91	0.91
Average	**1**	**0,88**	**0,88**	**0,87**	**0.81**

Results: Action Selection



CityLearn Challenge 2022

- Public Challenge to perform the best on a private CityLearn Dataset
- I did not have a good solution ready before the deadline
- I won the Community Contribution Prize for finding and fixing bugs, including a ~100x speed improvement on CityLearn’s simulation.

Discussion

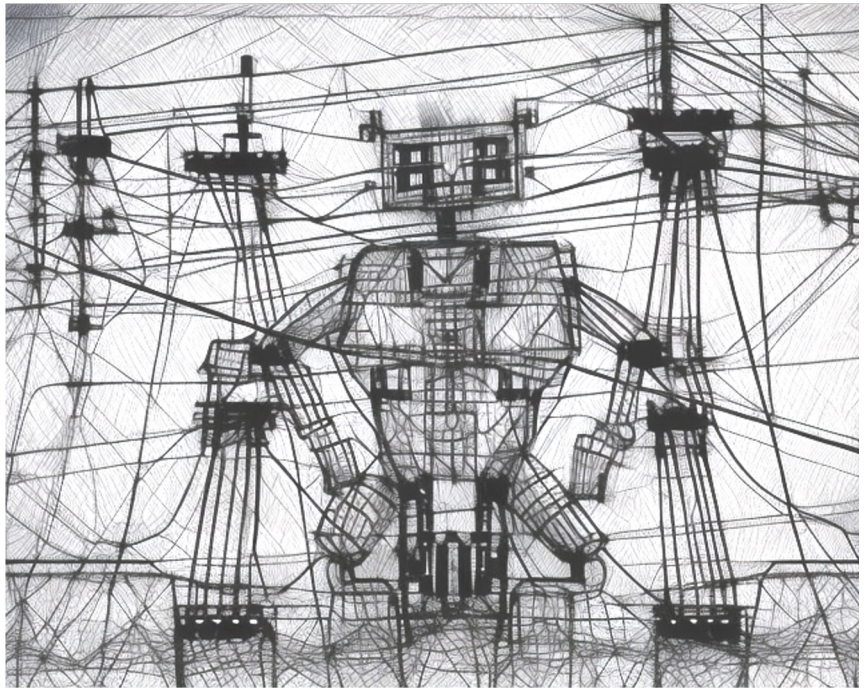
- UA-DQN is a suitable algorithm for Demand Response.
- It allows for more efficient action selection, and therefore **higher sample efficiency**
- It is still outclassed by my simple hand-engineered policy
- Reinforcement Learning is tricky to get right!
  - Increase Resource constraints?
  - Different model architecture?
  - Different initialization?
  - More features?

Further Research

- Test generalization to other buildings (I will do this)
- Test different risk-aware strategies (I plan to do this)
- Multi-Agent coordination
- Non-stationary environment

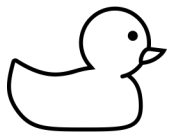
Summary: Key points

1. **Demand Response** is needed for flexibility in the Grid
2. In **Q-Learning**, efficient **action selection** is important
3. An explicit treatment of **uncertainty in RL** can improve
   - **data-efficiency**
   - **performance**
   - **robustness (?)**
   - **interpretability (?)**



References

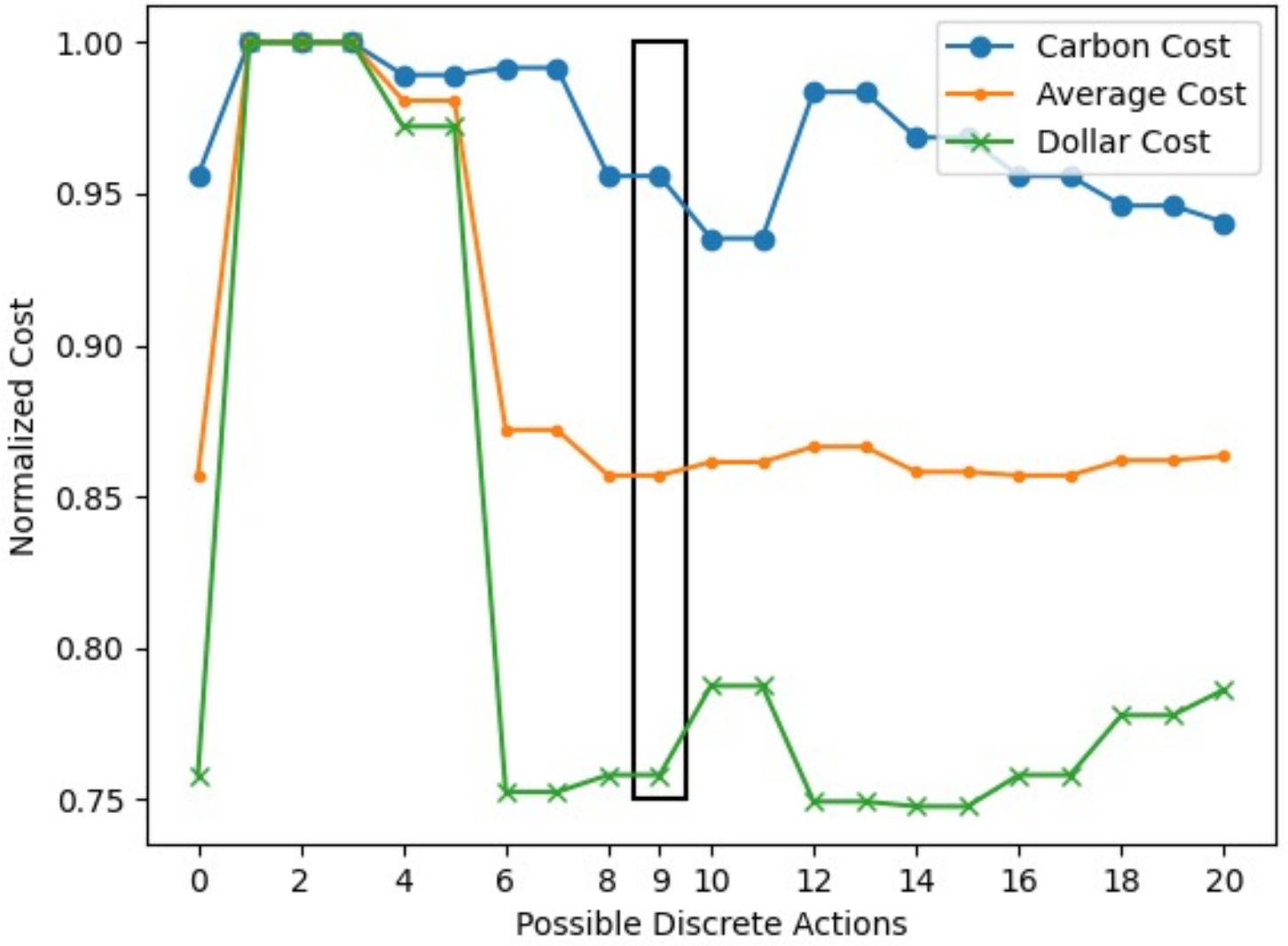
- Vázquez-Canteli, J. R., Kämpf, J., Henze, G., & Nagy, Z. (2019, November).
  CityLearn v1. 0: An OpenAI gym environment for demand response with deep reinforcement learning.
  In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 356-357)
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015).
  Human-level control through deep reinforcement learning.
  nature, 518(7540), 529-533.
- Clements, W. R., Van Delft, B., Robaglia, B. M., Slaoui, R. B., & Toth, S. (2019).
  Estimating risk and uncertainty in deep reinforcement learning.
  arXiv preprint arXiv:1905.09638.
- Illustrations: Stable Diffusion



Appendix

Discretization

- DQN and UA-DQN assume a discrete action space
- More actions = more complexity
- 9 preserves the middle action and has comparable performance to continuous steps



Notes on Data:

- 5 buildings, 1 year
- Hourly observations
- I use:
  - hour (1-hot encoded)
  - Carbon intensity
  - Electrical load
  - Solar Generation
  - Battery State of Charge
  - Pricing

```
0 month
1 day_type
2 hour
3 outdoor_dry_bulb_temperature
4 outdoor_dry_bulb_temperature_predicted_6h
5 outdoor_dry_bulb_temperature_predicted_12h
6 outdoor_dry_bulb_temperature_predicted_24h
7 outdoor_relative_humidity
8 outdoor_relative_humidity_predicted_6h
9 outdoor_relative_humidity_predicted_12h
10 outdoor_relative_humidity_predicted_24h
11 diffuse_solar_irradiance
12 diffuse_solar_irradiance_predicted_6h
13 diffuse_solar_irradiance_predicted_12h
14 diffuse_solar_irradiance_predicted_24h
15 direct_solar_irradiance
16 direct_solar_irradiance_predicted_6h
17 direct_solar_irradiance_predicted_12h
18 direct_solar_irradiance_predicted_24h
19 carbon_intensity
20 non_shiftable_load
21 solar_generation
22 electrical_storage_soc
23 net_electricity_consumption
24 electricity_pricing
25 electricity_pricing_predicted_6h
26 electricity_pricing_predicted_12h
27 electricity_pricing_predicted_24h
```

- Tuning Results
  best hyperparameters

	0	1	2
action_selection	egreedy	softmax	NaN
adam_epsilon	0.0	0.00001	0.0
final_exploration_rate	0.02	0.02	NaN
final_exploration_step	1000.0	1000.0	NaN
gamma	0.99	0.99	0.99
initial_exploration_rate	1.0	1.0	NaN
learning_rate	0.00007	0.0003	0.0003
loss	mse	mse	NaN
minibatch_size	4	128	128
replay_buffer_size	10000	10000	10000
replay_start_size	4	128	128
train_steps	438000	438000	438000
update_frequency	1	1	1
update_target_frequency	16	4	4
agent	DQN	DQN	UADQN
aleatoric_factor	NaN	NaN	0.0
biased_aleatoric	NaN	NaN	False
epistemic_factor	NaN	NaN	1.0
n_quantiles	NaN	NaN	20.0
noise_scale	NaN	NaN	1.0

Renewables need flexible backup, not baseload

Estimated power demand over a week in 2012 and 2020, Germany

Source: Volker Quaschning, HTW Berlin