SustainGym: Reinforcement Learning Environments for Sustainable Energy Systems

Christopher Yeh[1], Victor Li[1], Rajeev Datta[1], Julio Arroyo[1], Nicolas Christianson[1], Chi Zhang[2], Yize Chen[3], Mehdi Hosseini[4], Azarang Golmohammadi[4], Yuanyuan Shi, Yisong Yue[1], Adam Wierman[1]

[1]California Institute of Technology [2]University of California, San Diego [3]Hong Kong University of Science and Technology [4]Beyond Limits

chrisyeh96.github.io/sustaingym

Abstract

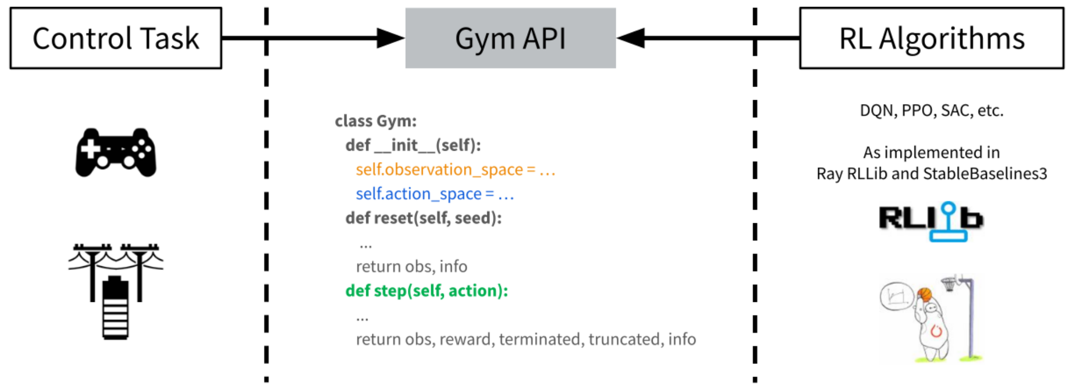
The lack of standardized benchmarks for reinforcement learning (RL) in sustainability applications has made it difficult to both track progress on specific domains and identify bottlenecks for researchers to focus their efforts. In this paper, we present SustainGym, a suite of five environments designed to test the performance of RL algorithms on realistic sustainable energy system tasks, ranging from electric vehicle charging to carbon-aware data center job scheduling. The environments test RL algorithms under realistic distribution shifts as well as in multi-agent settings. We show that standard off-the-shelf RL algorithms leave significant room for improving performance and highlight the challenges ahead for introducing RL to real-world sustainability tasks.

Background

A RL environment is a Markov Decision Process (MDP) which describes a specific control task and consists of:

- $\mathcal{S}$: state space
- $\mathcal{A}$: action space
- $\mathcal{P}$: transition kernel $\quad \mathcal{P}(s,r,a,s') = \Pr(s_{t+1} = s', r_{t+1} = r \mid s_t = s,\ a_t = a)$
- γ: discount factor

A "gym" is an abstraction in code for a MDP



No existing RL benchmarking environments specifically test multi-agent RL under distribution shifts.

Acknowledgments

We would like to acknowledge Steven Low, Tongxin Li, Zachary Lee, Lucien Werner, Zaiwei Chen, Ivan Jimenez, Pieter Van Santvliet, and Ameera Abdelaziz for their feedback and input during the preparation of SustainGym. The model underlying CogenEnv was developed in partnership with Beyond Limits and Enexsa. Funding for this project comes from a variety of sources, including NSF awards CNS-2146814, CPS-2136197, CNS-2106403, EPCN-2200692, and NGSDI-2105648; an NSF Graduate Research Fellowship; an Amazon AI4Science Fellowship; the Resnick Sustainability Institute; Hellman Fellowship; Amazon Web Services; and Beyond Limits.

This work is partially supported by the Department of the Air Force under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Air Force. Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

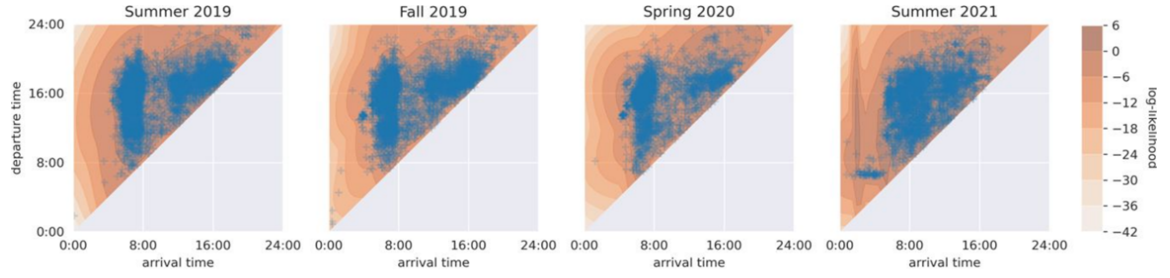
SustainGym Environments

	EVChargingEnv	ElectricityMarketEnv	DataCenterEnv	CogenEnv	BuildingEnv
Control task	charging rates for EV charging stations	market bids for a grid-connected battery storage system	virtual capacity curve for a carbon-aware data center	dispatch set points for turbines in a natural gas power plant	heating supply for rooms in a building
Modeled after	charging networks at Caltech & JPL	generic test case (IEEE RTS-GMLC)	(loosely) a Google data center	specific combined cycle gas generation plant in the U.S.	generic DoE commercial reference building models
Single agent	all EV charging stations	single battery system	single data center	all 4 turbine units	all buildings
Multi-agent	one EV charging station	N/A	N/A	one turbine unit	one room
Actions	discrete or continuous	discrete or continuous	continuous	mixed discrete & continuous	discrete or continuous
Rewards	energy charged - cost - CO_2 emissions	profit - CO_2 emissions	jobs scheduled - penalty + CO_2 emissions	- fuel consumption - constraint penalty	- temperature deviation - energy use
Distribution Shift	MOER, EV arrivals	MOER, load	MOER	renewable wind penetration	outdoor temperature

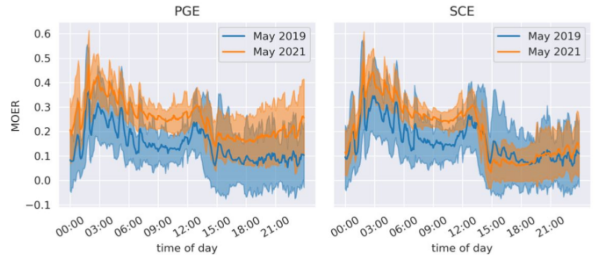
Distribution shifts are ubiquitous in real-world environments

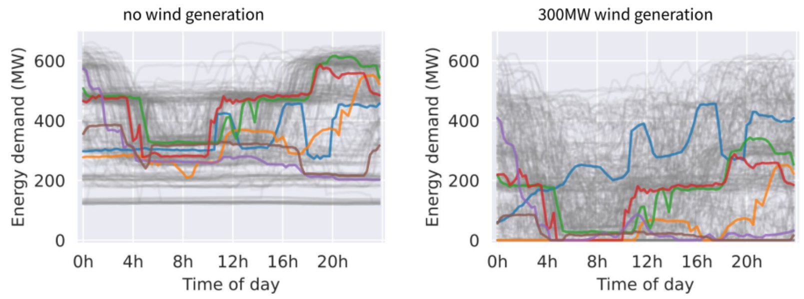
A distribution shift is a change in the transition kernel $\mathcal{P}$

Arrival/departure patterns of EVs changed significantly between pre-, during, and post-COVID



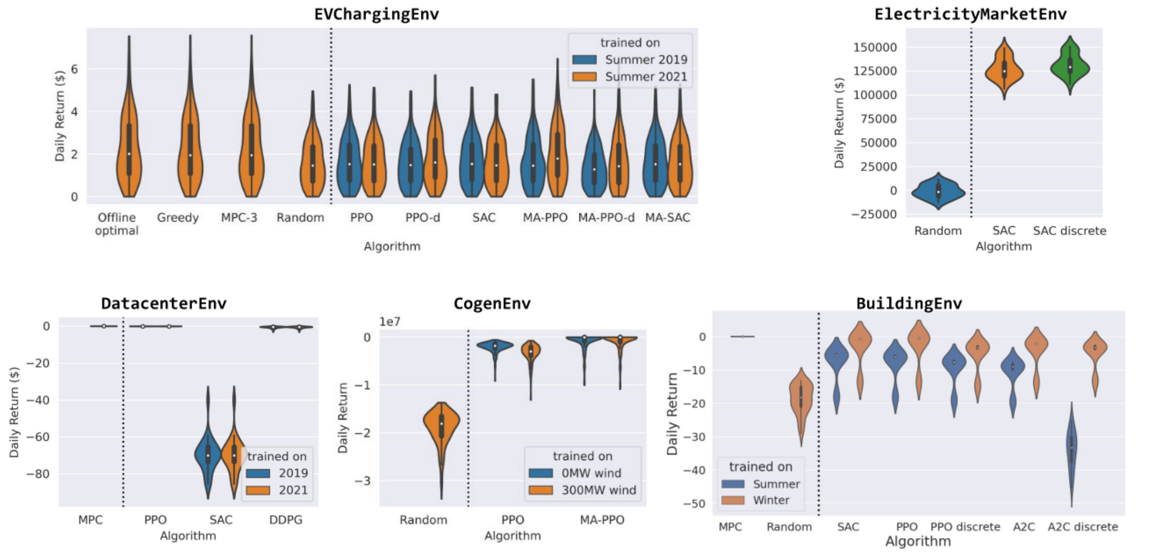
Carbon emissions rates (kg CO_2 / kWh) change over space and time





Net energy demand changes as renewable wind generation capacity increases over time

Benchmark Results



Key takeaways

- Off-the-shelf RL algorithms have significant room for improvement, especially compared to non-RL algorithms such as MPC
- Multi-agent RL generally performs on-par with single-agent RL
- No existing off-the-shelf RL algorithm handles distribution shift well

Future Directions

SustainGym serves as a test bed for **multi-agent model-free distributionally-robust RL**

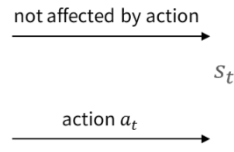
$$\max_{\pi} \min_{\mathcal{P} \in \mathbb{P}} \mathbb{E}_{\pi,\mathcal{P}} \left[\sum_{t=0}^{T} \gamma^t R(s_t, \pi(s_t)) \right]$$

where $\mathbb{P}$ denotes an uncertainty set of transition kernels

SustainGym motivates research into **RL algorithms that take advantage of knowledge of causal structure in the state space**

- For example, the only part of `EVChargingEnv` that is explicitly affected by the charging action taken is the electricity demand d_t remaining for each EV. The remainder of the observation evolves independently from the action taken.

m_t = carbon emissions rate
e_t = estimated time until departure
d_t = remaining energy demand



$$s_t = \begin{bmatrix} t \\ m_t \\ \hat{m}_{t:t+k-1|t} \\ e_t \\ d_t \end{bmatrix} \qquad s_{t+1} = \begin{bmatrix} t+1 \\ m_{t+1} \\ \hat{m}_{t:t+k-1|t} \\ e_{t+1} \\ d_{t+1} \end{bmatrix}$$

SustainGym motivates research into **multi-agent RL algorithms that can leverage network structure**

- For example, `BuildingEnv` features controllable AC units in different rooms, but only certain rooms share a wall, and heat transfer happens through walls