What Should We Grow Today so We Make Money Tomorrow

Reinforcement Learning for Small Farmers

Nick Lu, Jasmine Stefano

Mentor: Aviva Prins

Table of Contents

Challenge

Our aim has been to develop an **optimization-driven support system** to produce actionable and explainable **instruction** for farmers in a realistic, dynamic environment, providing a tool geared towards **maximizing profits**.

KHEYTI

Decision support system

Example policy:

- Plant **eggplant** in December 2021
- Plant **cucumber** in March 2022
- Plant **beetroot** in July 2022
- Plant **cucumber** in October 2022

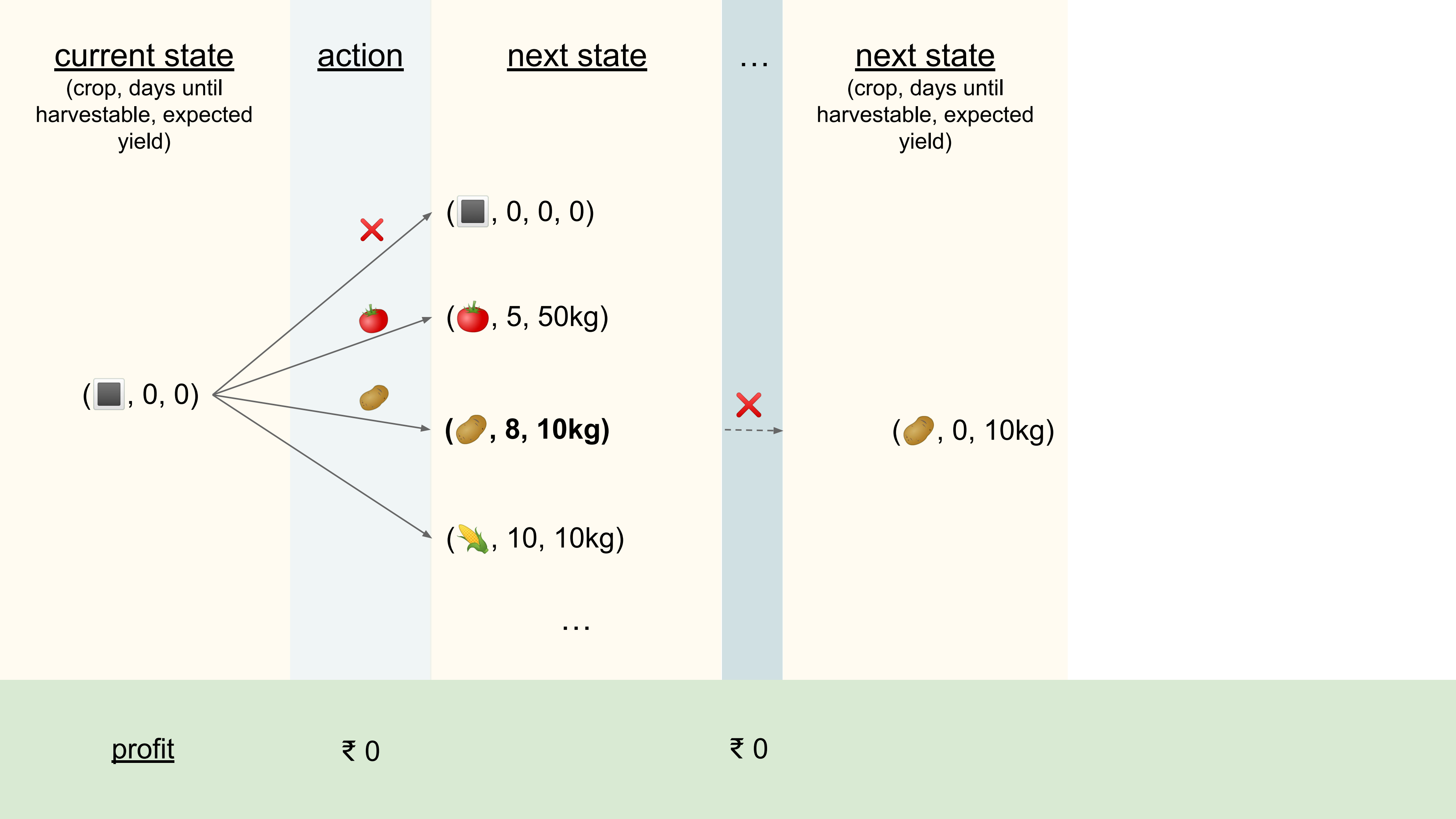
Markov Decision Processes (MDP)

State space: (crop, maturity, expiry, flag)

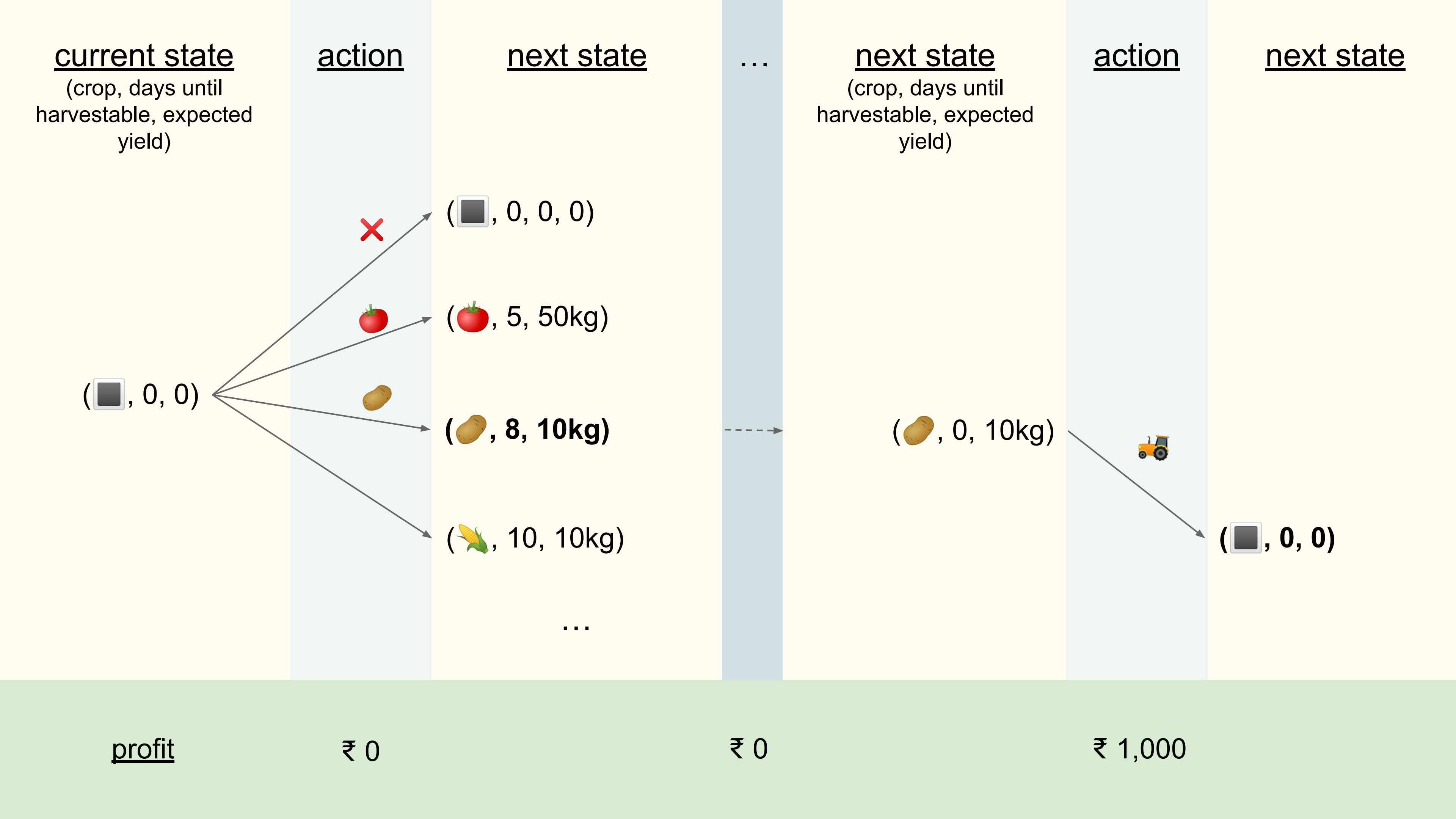
Action space: (no act, plant crop{1,...,n}, harvest)

Transition function: P(S,a,S’)

Reward function: $R(S,a) = \begin{cases} < 0 & \text{if action yields constraint violation} \\ \$(\text{crop}) & \text{if action is harvest} \\ 0 & \text{otherwise} \end{cases}$



current state
(crop, days until harvestable, expected yield)
action
next state
…
next state
(crop, days until harvestable, expected yield)
(⬛, 0, 0)
❌
🍅
🥔
(⬛, 0, 0, 0)
(🍅, 5, 50kg)
(🥔, 8, 10kg)
(🌽, 10, 10kg)
…
❌
(🥔, 0, 10kg)
profit
₹ 0
₹ 0


current state
(crop, days until harvestable, expected yield)
action
next state
…
next state
(crop, days until harvestable, expected yield)
action
next state
(⬛, 0, 0)
❌
(⬛, 0, 0, 0)
🍅
(🍅, 5, 50kg)
🥔
(🥔, 8, 10kg)
(🌽, 10, 10kg)
…
(🥔, 0, 10kg)
🚜
(⬛, 0, 0)
profit
₹ 0
₹ 0
₹ 1,000

Reinforcement Learning

Goal: Maximize expected total discounted reward:

$$\mathbb{E}[\sum_{t=0}^{H} \gamma^t R(s_t, a_t) | \pi]$$

H : Horizon

γ : discount factor

R : reward function

π : policy

Reinforcement Learning

Bellman Equation:

$$V(s) = \sum_{s'} P(s, \pi(s), s')(R(s, \pi(s)) + \gamma V(s'))$$

s : state $\quad s'$: next state

P : transition function

γ : discount factor

R : reward function

π : Policy

V : Value function

Reinforcement Learning

Choose action such that it yields the max expected reward:

$$\pi(s) = \arg\max_a \sum_{s'} \underbrace{P(s,a,s')}_{\text{transition}} (\underbrace{R(s,a)}_{\text{reward}} + \gamma \cdot \underbrace{V(s')}_{\text{future reward}})$$

s : state $\quad s'$: next state

a : action

γ : discount factor

V : Value function

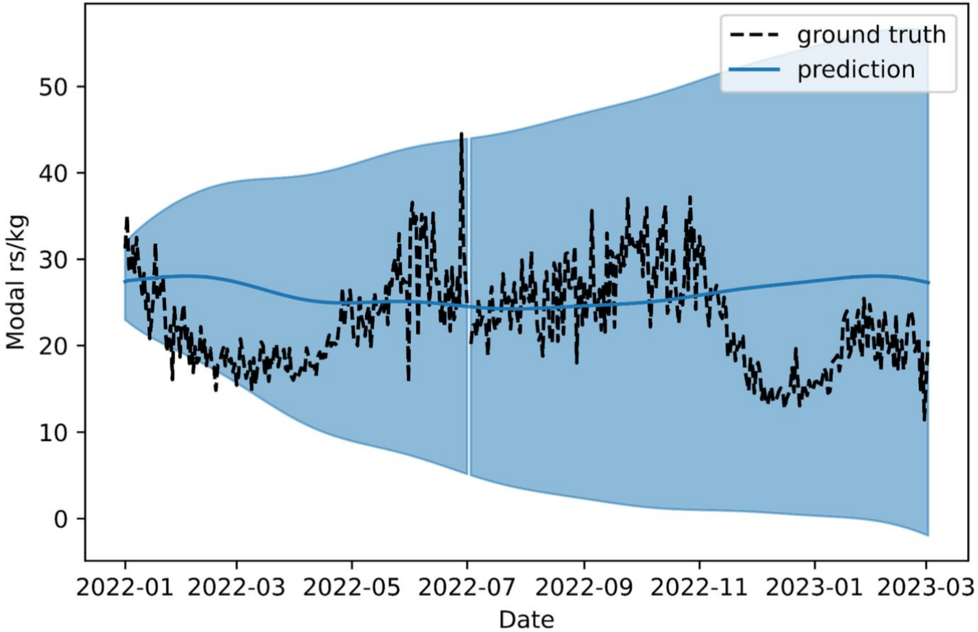
Offline vs. Online

Offline Reinforcement Learning

Reinforcement Learning with Online Interactions

Previous Work

The offline implementation uses forecast models to precompute **predicted future market prices.**



Online Algorithm

Algorithm 1 Follow the Weighted Leader for MDP

Input: Transition matrix P, parameter $\theta \in [0, 1)$, initial state s_0
Initialization: $\hat{R}_0$

1: **for** $t = 1 : H$ **do**
2: Update the weighted average of history rewards:

$$\hat{R}_t = (1 - \theta)\hat{R}_{t-1} + \theta R_{t-1}$$

3: Solve the MDP given reward matrix $\hat{R}_t$ for the average optimal policy:

$$\pi_t \in \arg\max_{\pi} g_{\hat{R}_t}(\pi)$$

4: Execute π_t, Update current State s_t
5: $R_t \leftarrow$ true reward matrix(from market data)
6: **end for**

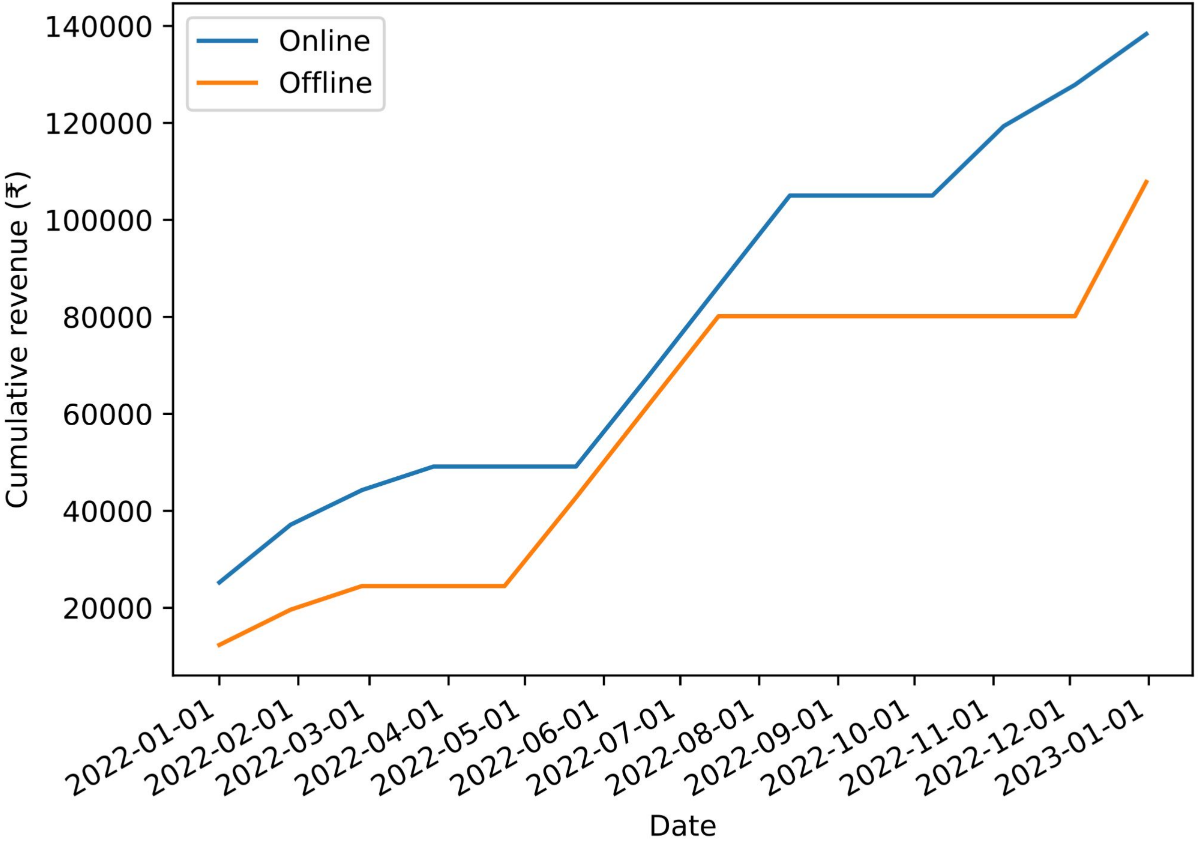
Output: π_t at each time step $t = 1, \ldots, H$

Y. Li and N. Li, "Online Learning for Markov Decision Processes in Nonstationary Environments: A Dynamic Regret Analysis", 2019

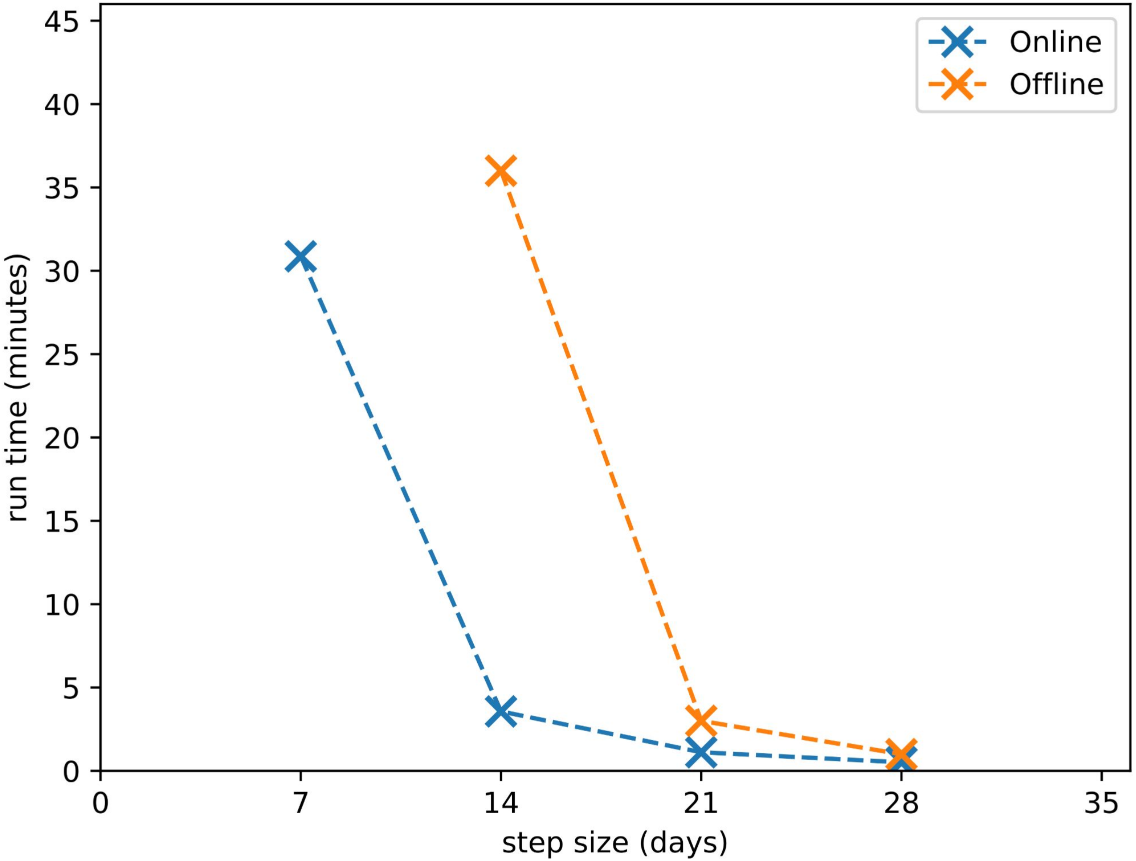
Benefits

- **Reduced state space complexity:**
  - By converting this model from offline to online, we took out the time component in state space, reducing the complexity of state space from order of 10^7 to 10^5, by a factor of 100.
- **Avoid inaccurate price forecast:**
  - Offline model suffers from inaccurate forecasting to generate policy
- **Smaller memory for storing data:**
  - Since online algorithm only needs current information, much smaller memory is required.

Simulation Results



Simulation Results



Thank you

References

- Y. Li and N. Li, "Online Learning for Markov Decision Processes in Nonstationary Environments: A Dynamic Regret Analysis," 2019 American Control Conference (ACC)

- What Should I Grow Today So I Make Money Tomorrow? Supporting Small Farmers' Crop Planning with Social, Environmental, and Market Data. A Prins, C Herlihy, JP Dickerson, Practical ML for Developing Countries Workshop, ICLR 2022