An Online Optimization-Based Decision Support Tool for Small Farmers in India: Learning in Non-stationary Environments

Tuxun Lu[1], Aviva Prins[2]

[1]Johns Hopkins University, [2]University of Maryland, College Park

[1]tlu32@jhu.edu, [2]aviva@cs.umd.edu

Introduction

- Small farmers in India are facing a decline in the productivity and profitability in the agricultural sector.
- Climate changes exacerbate the situation.
- Crop management decision support systems can help farmers to reduce riskiness of revenue stream, but they do not have access.

Objectives

- Address the challenges faced by small farmers in India by developing an online optimization-based decision support tool for crop management.
- The decision support tool should be cheap to deploy, adaptive to market changes, and resource efficient.

Methods

State space

$$S \coloneqq \{(crop, maturity, expiry, flag)\}$$

Action space

$$A \coloneqq \{no\ act, harvest\} \cup \{plant(c) \mid \forall c \in C\}$$

Transition function

$$P_t : S \times A \times S \to \{0,1\}$$

Reward function

$$R_{t(s,a)} \coloneqq \begin{cases} k, & \text{constraint violation} \\ y_t(s,a), & a = harvest \\ 0, & \text{otherwise} \end{cases}$$

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_t(s_t, a_t) \mid \pi\right]$$

$$V(s) = \max_a \sum_{s'} P(s,a,s')\left[R(s,a) + \gamma V(s')\right]$$

Algorithm

Algorithm 1: Offline version

Input: Historical price data, initial state s_0, transition matrix P

Forecast: Approximate $\hat{R}$ with a forecast of historical data.

1: Solve the MDP $(S, A, P, \hat{R})$ for a policy:

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

Output: $\pi = \{\pi_t \mid t = 1, \ldots, T\}$

$$|\mathcal{S}| = 2|\mathcal{C}| \times \max_{c \in \mathcal{C}}(\texttt{c.max\_maturity}) \times \max_{c \in \mathcal{C}}(\texttt{c.lifespan})$$

# of Entries	Online	Offline
State Space	560	97,440
Transition	3,136,000	94,945,536,000

Algorithm 2: FWL with time-varying transition function P_t

Input: Smoothing parameter $\theta \in [0,1)$, initial state s_0, transition matrices $\{P_t\}$

Initialization: $\hat{R}_0 \leftarrow R_{-1}$

1: **for** $t = 1 : T$ **do**

2: Update the weighted average of historical rewards:

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

3: Solve the MDP $(S, A, P_t, \hat{R}_t)$ for a policy:

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

4: Execute π_t to transition from s_{t-1} to s_t

5: **end for**

Output: π_t at each timestep $t \in \{1, \ldots, T\}$

Li, Yingying, and Na Li. "Online learning for markov decision processes in nonstationary environments: A dynamic regret analysis." 2019 American Control Conference (ACC). IEEE, 2019.

Results

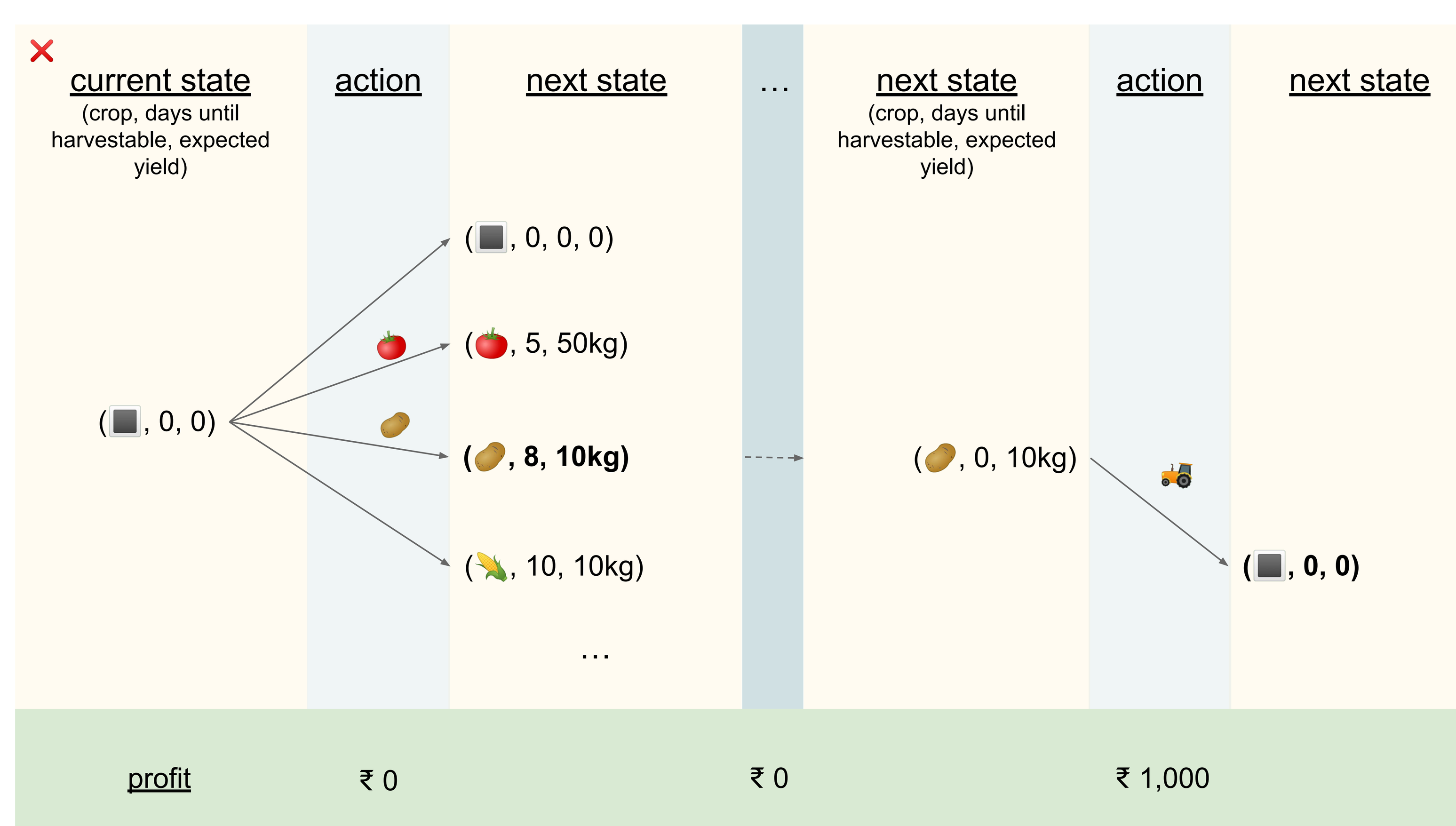

Figure 1: Illustration of state progression under different actions

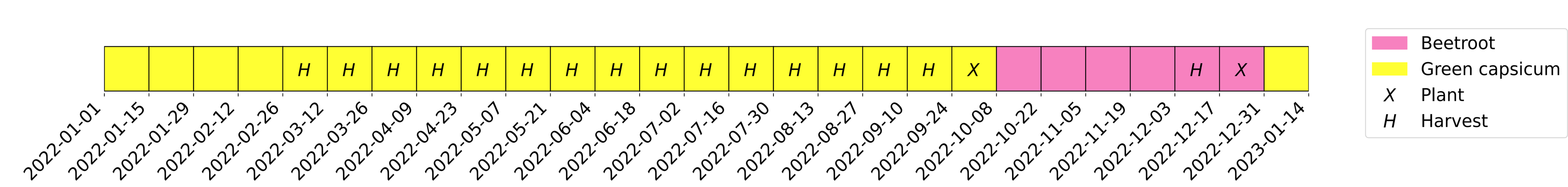

Figure 2: One progression of states and actions yielded by our simulations with high revenue

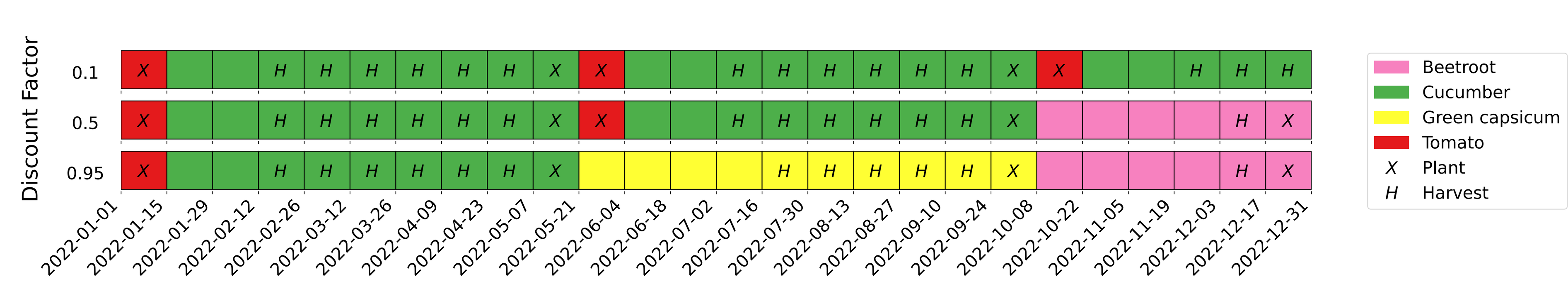

Figure 3: Progression of states and actions yielded by changing the discount factor γ

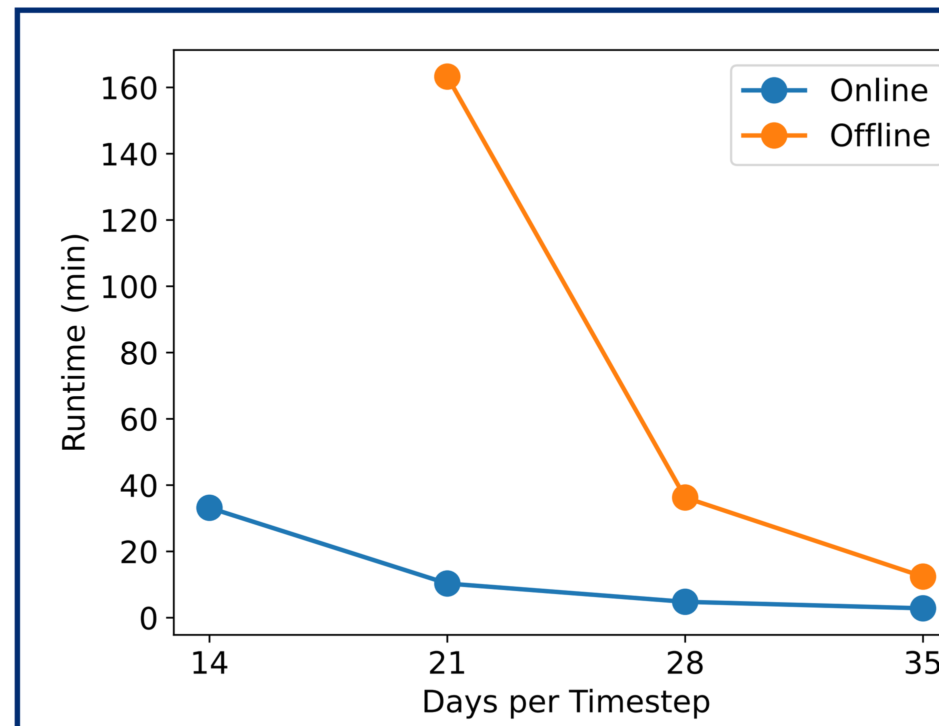

Figure 4: Comparison of runtimes over different timestep sizes (the simulation length remains ten years).

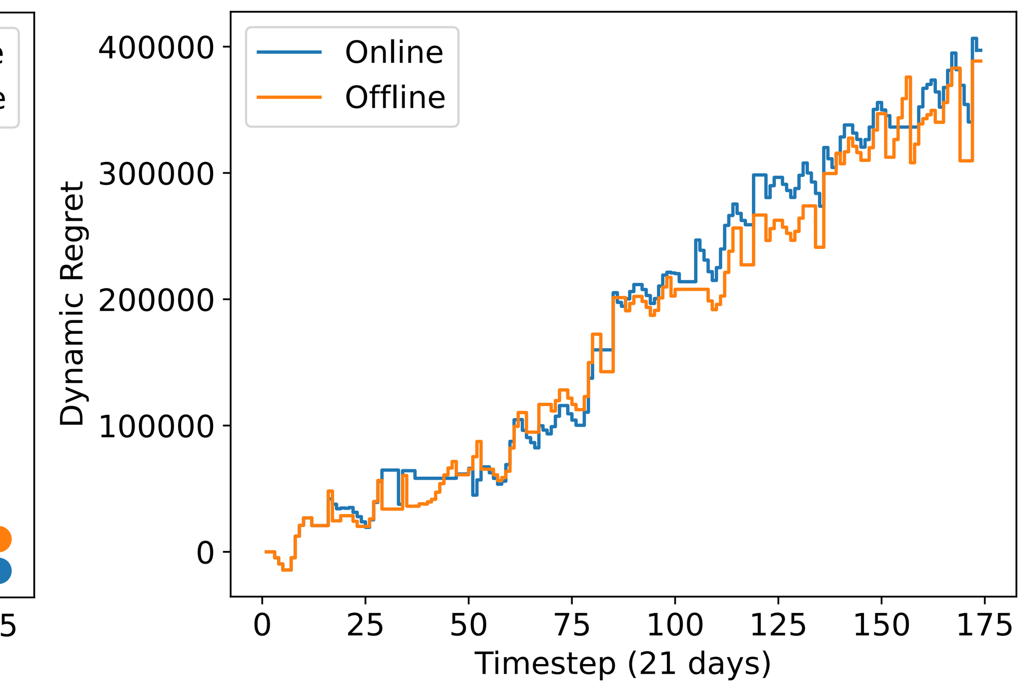

Figure 5: The dynamic regret of the online and offline outputs

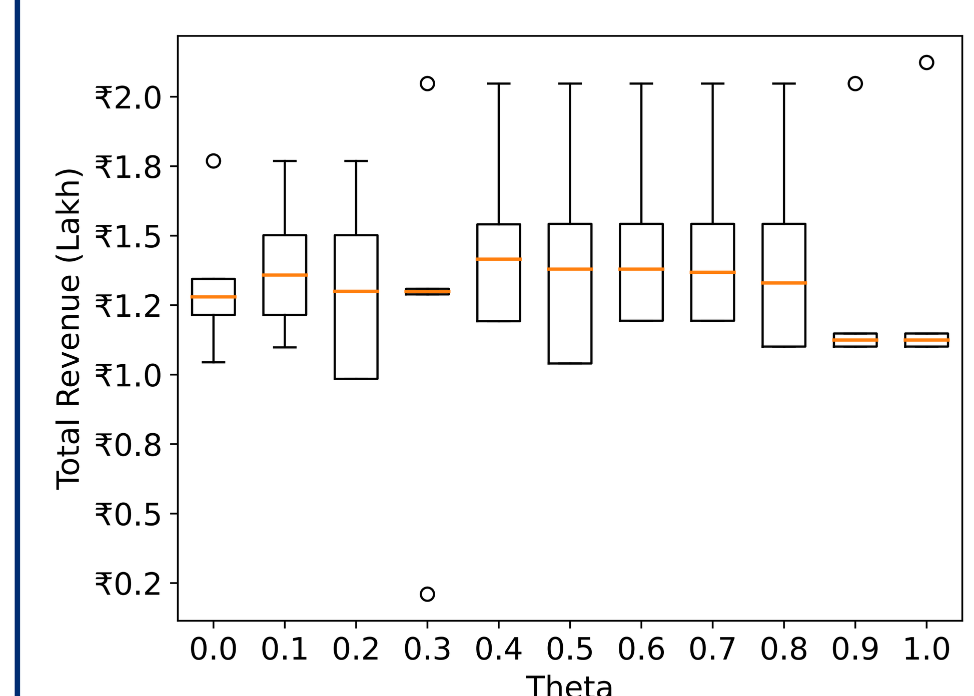

Figure 6: Cumulative revenue yielded by each policy over multiple values of θ

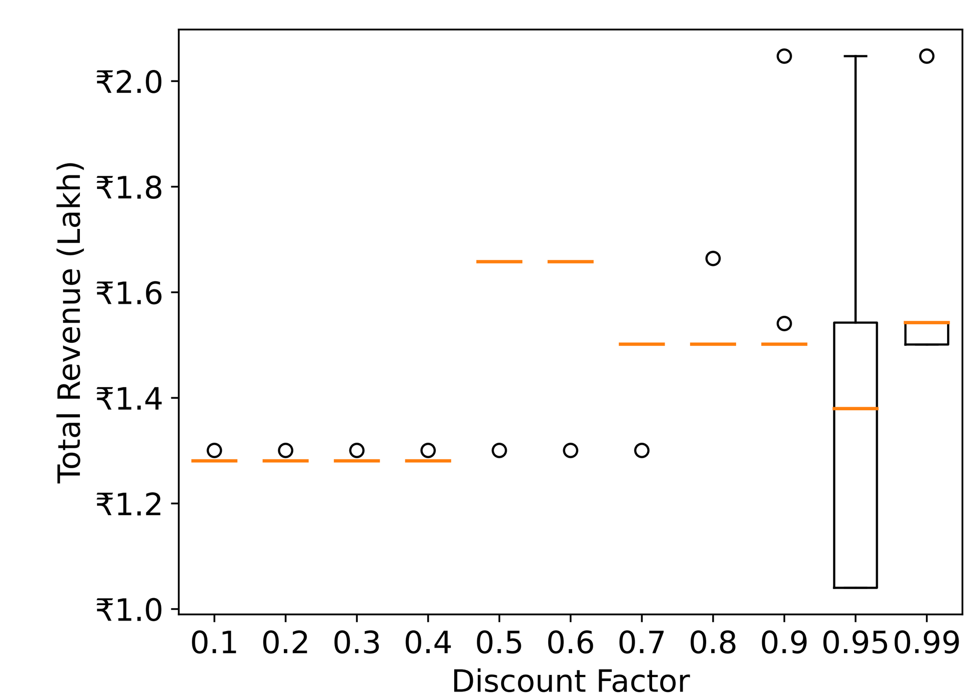

Figure 7: Immediate versus delayed reward prioritization yields different policies and cumulative revenues.

Conclusions & Future Work

- The policy of actions produced by FWL are fast to generate, intuitive, produce high cumulative revenues, and are approximately equivalent to a planning alternative.
- In the future, we aim to study the problem in a multi-agent setting. The transition function could be probabilistic to reflect real world crop growing. We are also interested in adjusting R to better reflect different goals, such as risk reduction, portfolio diversification.

Resources

Arxiv: https://arxiv.org/pdf/2311.17277.pdf

Acknowledgements

We thank Christine Herlihy and Jasmine Stephano for helpful research discussions and code contributions. Many thanks to Dileep K H, Kaushik Kappagantulu, and the team at Kheyti for partnership and feedback, especially in developing the Markovian model. This project was partially funded by the NSF REU-CAAR grant 2150382 and NSF CAREER Award IIS-1846237. The support of JHU WSE Undergraduate Conference Travel Fund and reimbursement from UMD enabled Tuxun Lu to attend this conference.