Decision-Making Under Selective Labels

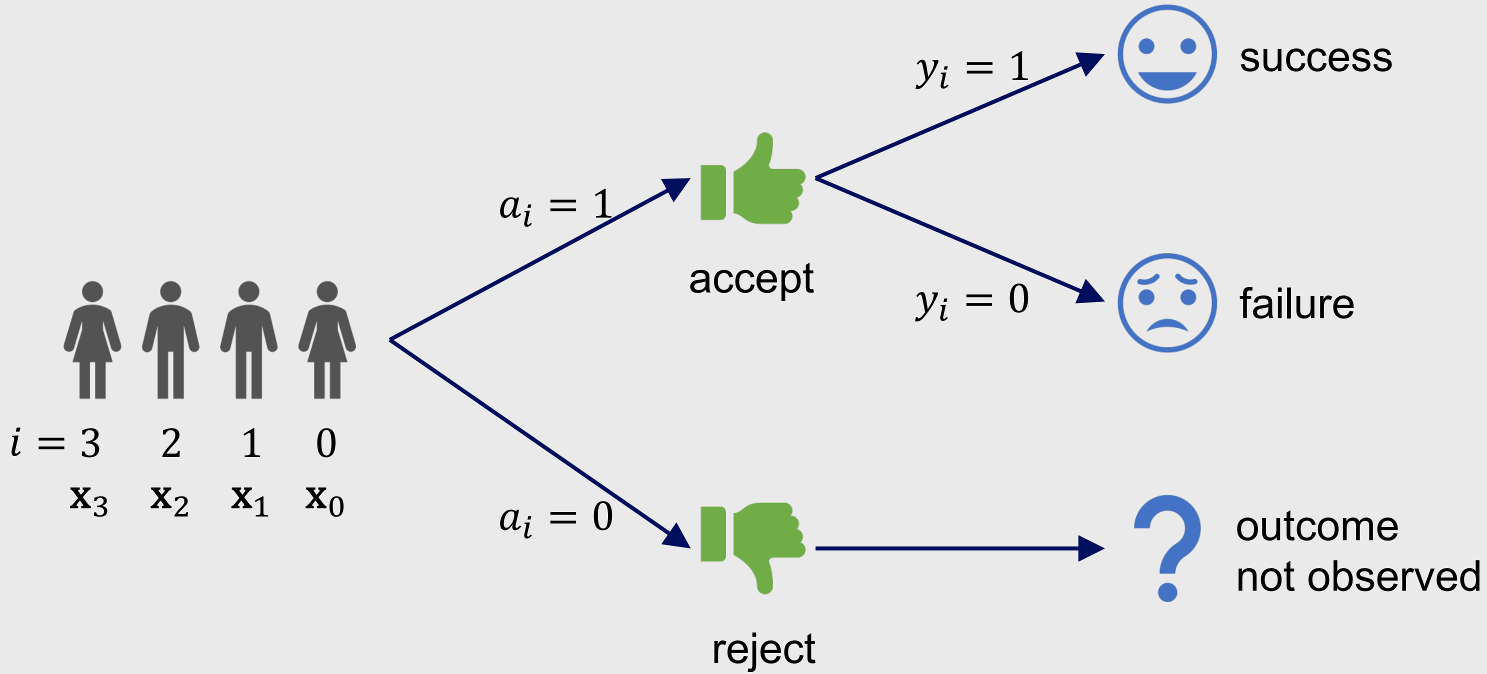
Optimal Finite-Domain Policies and Beyond

—

Dennis Wei
IBM Research

Selective Labels

Learn to make decisions with **no observed outcomes** under one of the decisions

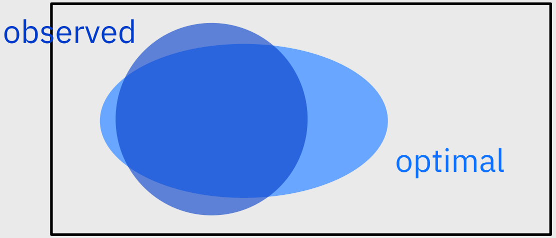


Existing Approaches

1. Supervised Learning

- Most common
- Threshold model predictions
- Update model based only on accepted individuals

Drawback: May be suboptimal due to censoring



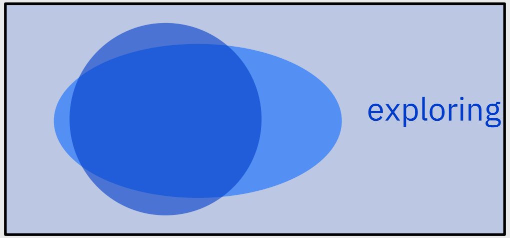
2. "Consequential Learning"

[Kilbertus et al., AISTATS 2020]

- Collect labelled data using existing policy
- Learn new policy to maximize held-out utility

Drawback 1: Needs labelled data from "exploring" policy

Drawback 2: Does not account for cost of this exploration



Proposed Online Formulation

Balance costs of decisions during learning against future utility

Learn decision policy $\Pi(x) = \Pr(A = 1|x)$ to maximize discounted total reward:

$$\mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i a_i(y_i - c)\right], \qquad \gamma < 1$$

$$a_i(y_i - c) = \begin{cases} 1 - c & \text{if success} \\ -c & \text{if failure} \\ 0 & \text{if reject} \end{cases}$$

Existing Approaches

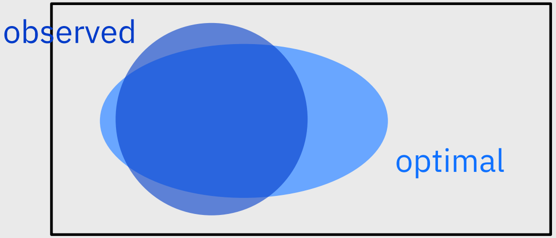
1. Supervised Learning

— Most common

— Threshold model predictions

— Update model based only on accepted individuals

Drawback: May be suboptimal due to censoring



2. "Consequential Learning"

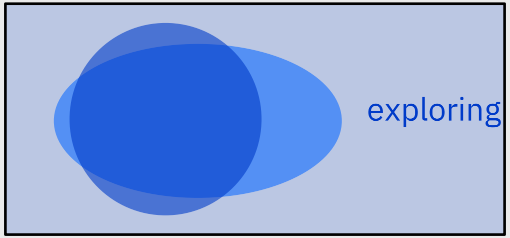
[Kilbertus et al., AISTATS 2020]

— Collect labelled data using existing policy

— Learn new policy to maximize held-out utility

Drawback 1: Needs labelled data from "exploring" policy

Drawback 2: Does not account for cost of this exploration



3. Contextual Bandits

— Two arms: accept/reject

— Context $\mathbf{x}$

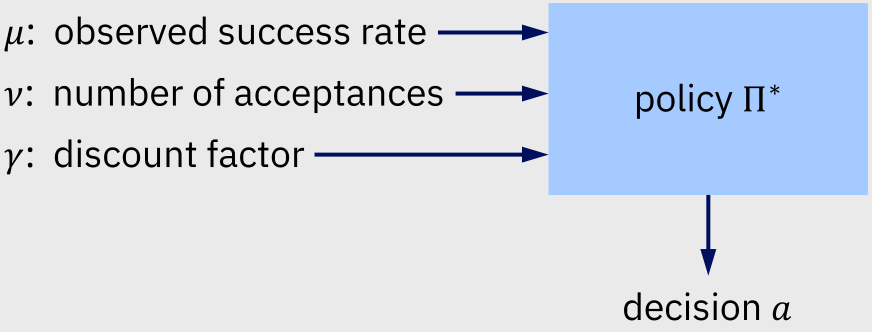
Drawback: Lower utility, due to not being tailored to selective labels problem

Approach: Start simple and generalize

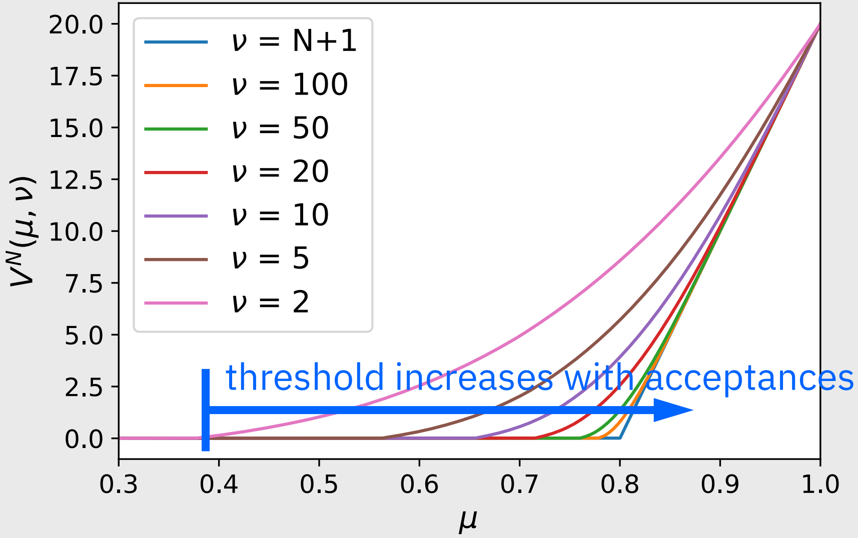
Homogeneous Case

Fix/drop X to give a *homogeneous* population

Dynamic programming yields **optimal** policy

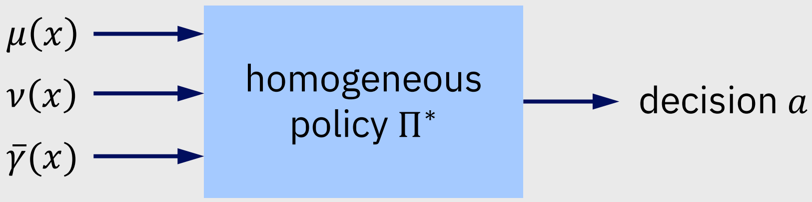


Deterministic: $\Pi^*(\mu, \nu) = \mathbf{1}(V^*(\mu, \nu) > 0)$



More General Cases

Leverage homogeneous policy



1. Finite Domain $X \in \mathcal{X},\ |\mathcal{X}| < \infty$

Optimal policy: homogeneous policies for $x \in \mathcal{X}$

$\mu(x), \nu(x)$: conditioned on x

$\bar{\gamma}(x)$: *effective* discount factor

2. Infinite Domain

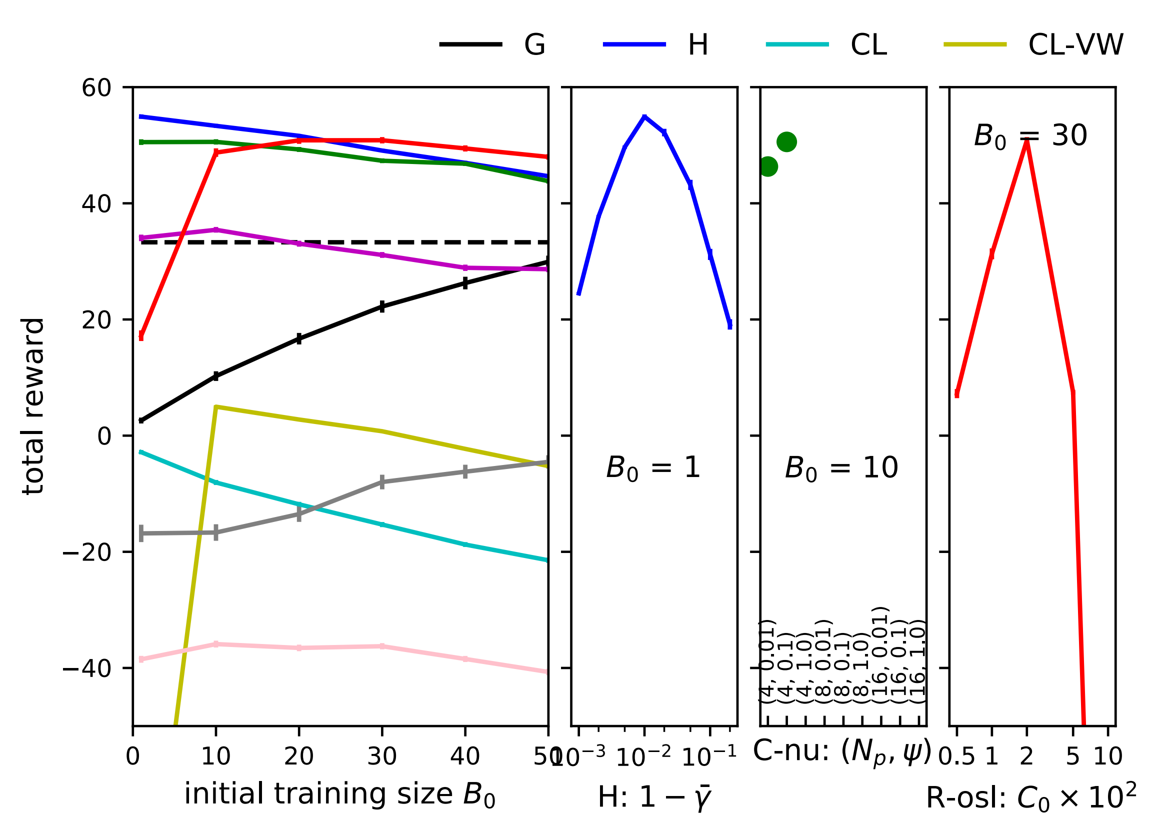
$\mu(x)$: success probability model

$\nu(x)$: confidence in $\mu(x)$ (using bootstrap)

$\bar{\gamma}(x) \equiv \bar{\gamma}$: exploration/exploitation parameter

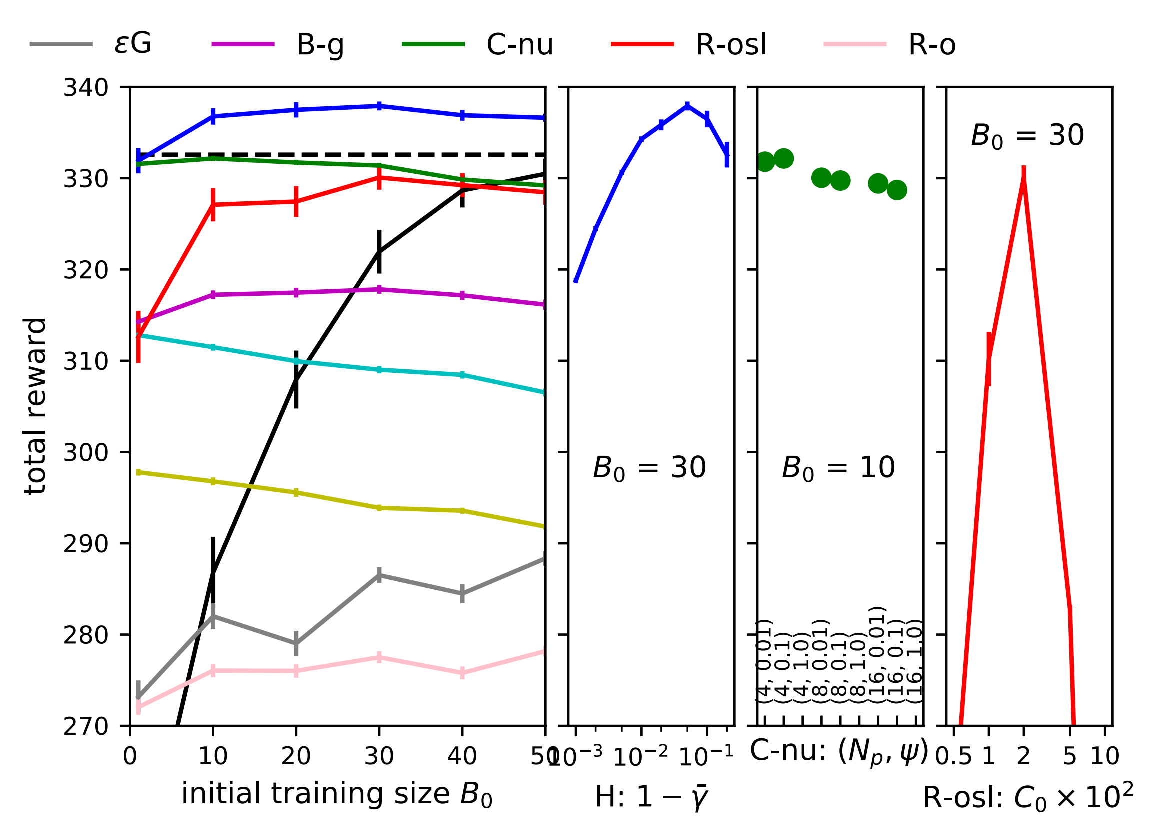
Utility on FICO Dataset (lending)

Proposed homogeneous policy in blue



R-osl (red) is a UCB-type policy adapted for selective labels

Utility on COMPAS Dataset (criminal justice)



Learn more at poster session: https://icml.cc/virtual/2021/poster/10109

Earlier version: https://arxiv.org/abs/2011.01381

© 2021 IBM Corporation