Counterfactual Structural Causal Bandits

Min Woo Park Sanghack Lee **(Seoul National University)**

Seoul National University
Graduate School of Data Science

ICLR
International Conference On Learning Representations

Multi-Armed Bandits

Multi-armed bandit (MAB) problem is a classic sequential decision-making problem.

Actions	a set of actions **A** to play; each action associates with a reward distribution.
Play	an action $A_{\mathbf{x}}$ for each round $t \in [T]$,
Reward	a reward $Y_{\mathbf{x}}$ is drawn from the reward distribution,
Goal	to minimize a cumulative regret over a total round T.

Structural Causal Bandits

- Multi-armed bandit through Causal Lens.
- A Structural Causal Model (SCM) $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$:
  - $\mathbf{U}$ unobserved variables;
  - $\mathbf{V}$ observed variables;
  - $\mathcal{F}$ causal mechanisms for $\mathbf{V}$ using $\mathbf{U}$ and $\mathbf{V}$;
  - $P(\mathbf{U})$ a joint distribution over $\mathbf{U}$ (**randomness**).
- Structural Causal Bandits: an MAB $\mathcal{M}$; a reward variable $Y \in \mathbf{V}$.

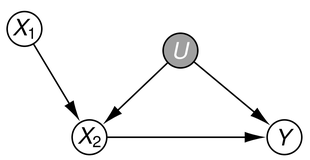
  Intervention sets (ISs) correspond to *all* subset of $\mathbf{V} \setminus \{Y\}$.

  Actions **A** correspond to values for intervention sets.
  i.e., action space $\{A_{\mathbf{x}} \mid \mathbf{x} \in \mathcal{D}_{\mathbf{X}}, \mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}\}$.

  Reward distribution $P(Y_{\mathbf{x}}) := P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$.

  Mean reward $\mu_{\mathbf{x}} := \mathbb{E}[Y \mid do(\mathbf{x})]$.
- **Assumption**. a causal diagram $\mathcal{G}$ (environment class) of $\mathcal{M}$ is accessible.

Example. We can control 2 binary variables, X_1 and X_2.



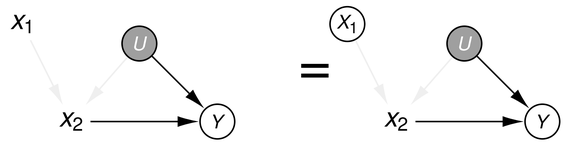
$$\mathcal{M} = \begin{cases} \mathbf{U} = \{U_1, U_2\} \\ \mathbf{V} = \{X_1, X_2, Y\} \\ \mathcal{F} = \begin{cases} f_{X_1}(u_1) = u_1 \\ f_{X_2}(x_1, u_2) = x_1 \oplus u_2 \\ f_Y(x_2, u_2) = x_2 \oplus u_2 \end{cases} \\ P(\mathbf{U}) = \{\forall U \sim \texttt{Bern}(0.5)\}. \end{cases}$$

Intervention sets (ISs) (**4**): $\emptyset$, $\{X_1\}$, $\{X_2\}$, $\{X_1, X_2\}$.

Actions (**9**): $\emptyset$, $do(x_1 = 0)$, $do(x_1 = 1)$, $\cdots$, $do(x_1 = 1, x_2 = 1)$.

Structural Properties

Property 1 (Equivalence among Actions)**.** Two actions share the same reward distribution.



$$\mu_{X_1, X_2} = \mu_{X_2}$$

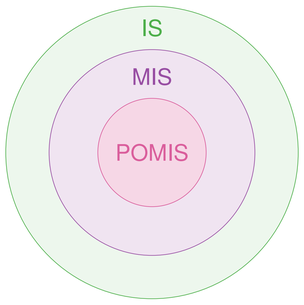
- **Minimal Intervention Set** (**MIS**): A minimal and ancestral set of variables among ISs sharing the same reward distribution.

Property 2 (Partial-Orderedness)**.** Maximum achievable expected rewards can be ordered.

$$\mu_{\emptyset} = \sum_{x_2} \mu_{x_2} P(x_2) \leq \sum_{x_2} \mu_{x_2^*} P(x_2) = \mu_{x_2^*}.$$

- **Possibly-Optimal Minimal Intervention Set** (**POMIS**): An MIS that can achieve an optimal expected reward in some SCM $\mathcal{M}$ conforming to the causal diagram $\mathcal{G}$ is called a POMIS.

Inclusion: All ISs $\supseteq$ MIS $\supseteq$ POMIS.

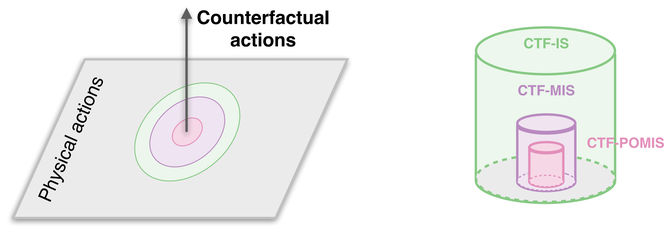


Cumulative Regrets: All ISs $\geq$ MIS $\geq$ POMIS (smaller the better).

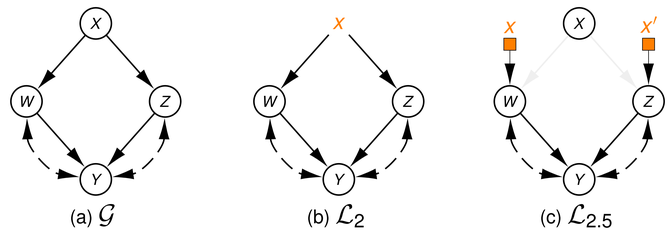
$\therefore$ A causal diagram allows an agent to optimize the action space *a priori*.

Extension to Counterfactual Action

- The action space is defined over $\mathcal{L}_1$ (*observational*) and $\mathcal{L}_2$ (*interventional*) of Pearl's Causal Hierarchy (PCH).
- We extend the notion of action into $\mathcal{L}_{2.5}$ (*realizable counterfactual*) (Raghavan and Bareinboim, 2025; Yang and Bareinboim, 2025)



Intuition. An action: a **node** intervention $\Rightarrow$ an **edge** intervention



For example, $do(W_x, Z_{x'})$ is a counterfactual action.

$$\begin{aligned} \mu_x &= \mathbb{E}[Y \mid do(x)] \\ &= \mathbb{E}[Y \mid do(W_x, Z_x)] \\ &\leq \mathbb{E}[Y \mid do(W_x, Z_{x'})] = \mu_{W_x, Z_{x'}} \end{aligned}$$

- We provide a sound and complete algorithm for enumerating all counterfactual-level POMISes (**CTF-POMIS**).

Conclusion

1. We generalize the notion of actions from the **physical** to the **counterfactual**.
2. We provide an efficient algorithm that identifies and eliminates redundant (equivalent) or suboptimal (partially ordered) actions.

Reference

Yang and Bareinboim, *A Hierarchy of Graphical Models for Counterfactual Inferences*, NeurIPS 2025

Raghavan and Bareinboim, *Counterfactual Realizability*, ICLR 2025

Lee and Bareinboim, *Structural Causal Bandits: Where to Intervene?*, NeurIPS 2018

Acknowledgements: This work was partly supported by the IITP (RS-2022-II220953/30%) and NRF (RS-2023-00211904/70%) grant funded by the Korean government.