Online Conformal Prediction with Adversarial Semi-Bandit Feedback via Regret Minimization

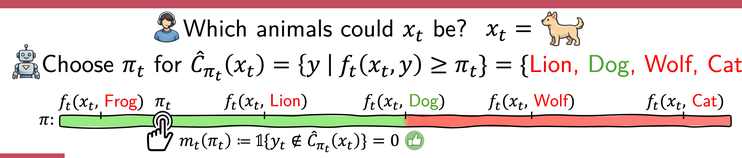
Junyoung Yang*[1] Kyungmin Kim*[2] Sangdon Park[12] (*: Equal contribution)
[1]POSTECH CSE [2]POSTECH GSAI

ICLR International Conference On Learning Representations

TL;DR

We introduce a novel algorithm for **online conformal prediction under adversarial semi-bandit feedback** that provably controls the **long-run miscoverage rate**.

Problem: Online Conformal Prediction Under Adversarial Partial Feedback

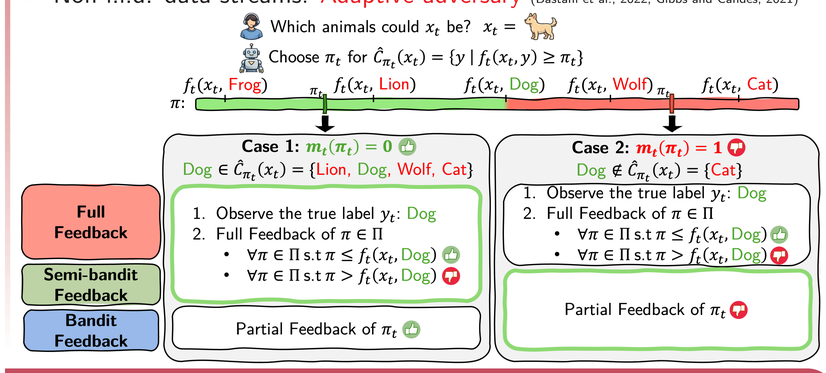


Goal: A certified long-run miscoverage rate control in online conformal prediction (OCP), while keeping conformal sets small

$$\mathbf{MC}(T) := \frac{1}{T}\sum_{t=1}^{T} m_t(\pi_t) \approx \alpha$$

Challenges in Real-world Scenarios:

- Limited amount of feedback: Semi-bandit feedback (Ge et al., 2025)
- Non-i.i.d. data streams: Adaptive adversary (Bastani et al., 2022; Gibbs and Candès, 2021)



Core Idea: Reformulating OCP as a Bandit Problem

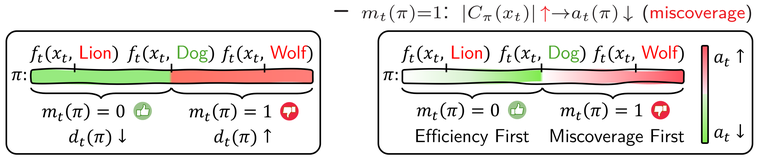
- Formulate OCP as a multi-armed bandit problem by treating each candidate conformal set as an arm
  - Discretize the hypothesis space $[0, 1]$ into Π, where $K := |\Pi|$

Problem	OCP w/ Adversarial Feedback	Multi-armed Adversarial Bandit
Adversary Controls	Sample: (x_t, y_t), where y_t is unlabeled	Loss Vector: $(\ell_t(\pi))_{\pi \in \Pi}$
Feedback (C1)	Miscoverage: $m_t(\pi_t)$ True Label: y_t if $m_t(\pi_t) = 0$	Loss: $\ell_t(\pi_t)$
Objective (C2)	Miscoverage Gap: $\mathbf{MC}(T) - \alpha$	$\mathbf{Reg}(T) := \sum_{t=1}^{T} \ell_t(\pi_t) - \min_{\pi \in \Pi} \sum_{t=1}^{T} \ell_t(\pi)$

Challenge 1 (C1): Design of the Loss Function

$$\ell_t(\pi) = \underbrace{d_t(\pi)}_{\text{miscoverage term}} + \underbrace{\lambda a_t(\pi)}_{\text{inefficiency term}}$$

- Miscoverage-first: $\ell_t(\pi_1) \leq \ell_t(\pi_2)$ $\forall \pi_1, \pi_2 \in \Pi$ s.t. $m_t(\pi_1) = 0$ and $m_t(\pi_2) = 1$
  - Strengthened for smaller α
- Why inefficiency term $a_t(\pi)$?
  - $m_t(\pi) = 0$: $|C_\pi(x_t)| \downarrow \to a_t(\pi) \downarrow$ (set size)
  - $m_t(\pi) = 1$: $|C_\pi(x_t)| \uparrow \to a_t(\pi) \downarrow$ (miscoverage)



Challenge 2 (C2): Regret-to-Miscoverage Gap Conversion

Algorithm-Agnostic Guarantees for Miscoverage Control

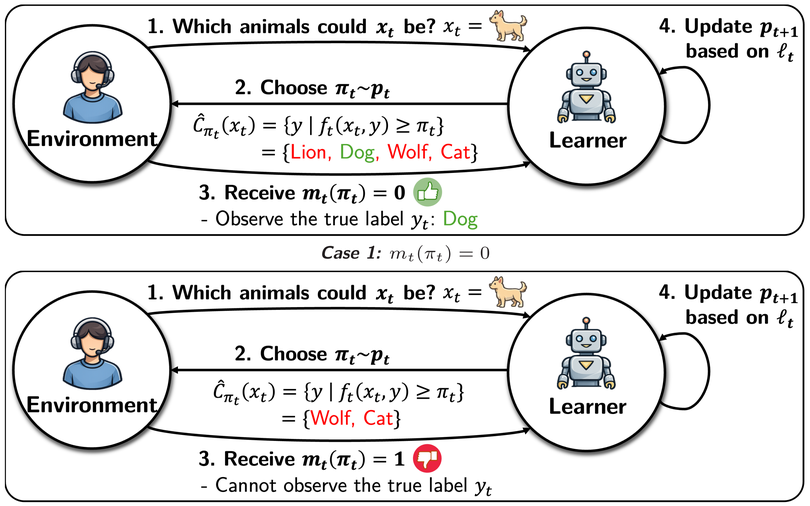
The deviation of the long-run miscoverage $\mathbf{MC}(T)$ from the target level α is bounded by the regret with respect to our proposed loss function:

$$\mathbf{MC}(T) - \alpha \leq \frac{1}{T}\mathbf{Reg}(T) + C_{\mathbf{MC}}(T),$$

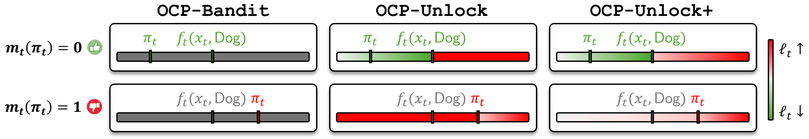
where $C_{\mathbf{MC}}(T)$ is a remainder term that vanishes as $T \to \infty$.

- Roughly speaking, any learner with sublinear regret controls the long-run miscoverage below α.
- We adapt EXP3.P algorithm which provides a high-probability sublinear regret even under an adaptive adversary.

Main Algorithm: OCP-Unlock+



[Overview of EXP3.P variants for OCP under adversarial semi-bandit feedback]



[Comparison of proposed EXP3.P variants with different update rules]

Guarantees on Miscoverage for OCP-Unlock+

For arbitrary data streams under an adaptive adversary, OCP-Unlock+ satisfies the following guarantee with high probability:

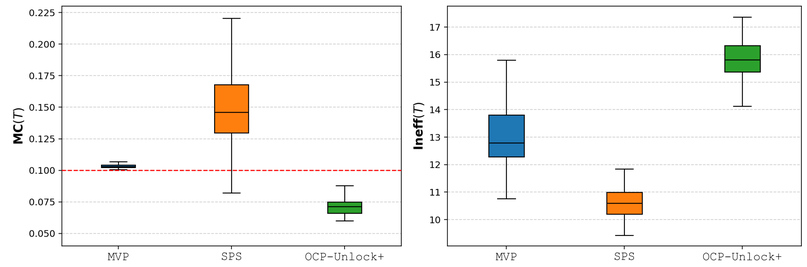
$$\mathbf{MC}(T) - \alpha \leq O\left(\sqrt{\frac{K \ln K}{T}}\right) + C_{\mathbf{MC}}(T).$$

Experiments & Results

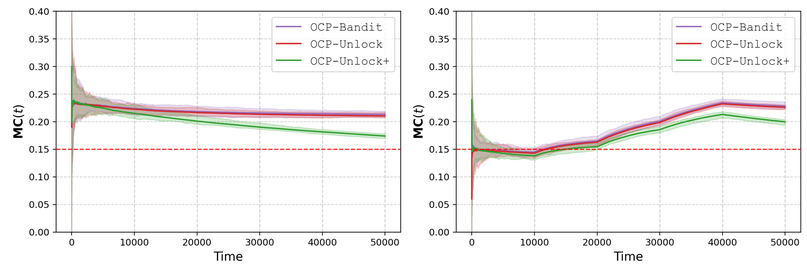
Evaluation Metric:

- $\mathbf{MC}(T) = \frac{1}{T}\sum_{t=1}^{T} m_t(\pi_t)$
- $\mathbf{Ineff}(T) = \frac{1}{T}\sum_{t=1}^{T} |\hat{C}_{\pi_t}(x_t)|$

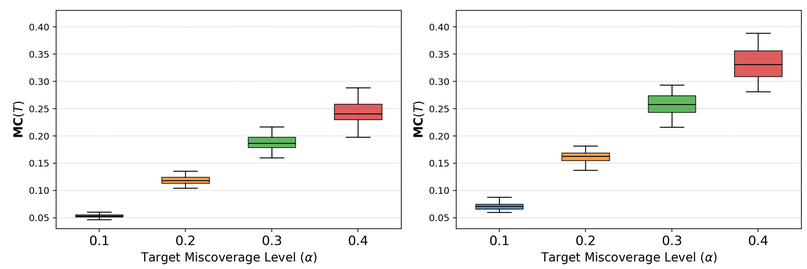
1. OCP-Unlock+ controls $\mathbf{MC}(T)$ even under non-i.i.d. data streams, comparable to an oracle baseline (MVP) with full feedback.



2. Fully exploiting semi-bandit feedback while respecting the monotonicity of miscoverage feedback with respect to the threshold is crucial for miscoverage control (Figure 4).



3. OCP-Unlock+ is stable across different α.



What Else You Can Find in Our Paper

- A detailed description of the loss function design
- A detailed algorithm description and comparison of the proposed EXP3.P variants
- Incorporation of the monotonicity of miscoverage feedback with respect to the threshold into the design of OCP-Unlock+