A Blackbox Approach to Best of Both Worlds in Bandits and Beyond

Presenter: Shinji Ito (NEC Corporation)

Authors: Christoph Dann (Google), Chen-Yu Wei (MIT), Julian Zimmert (Google)

The Best of Both (Three) Worlds Problem

(Figures taken from Wouter Koolen’s slides)

World	Stochastic	Corrupted	Adversarial
Regret bound	$\mathcal{O}(\log T)$	$\mathcal{O}(\log T + \sqrt{C \log T})$	$\mathcal{O}(\sqrt{T})$

(omitting other problem-dependent constants)

Goal: A single algorithm that has all guarantees without knowing the type of the world?

Existing Techniques for Multi-Armed Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound

Refined gap bound means obtaining $\sum_{i=1}^{K} \frac{\log T}{\Delta_i}$ instead of $\frac{K \log T}{\Delta_{\min}}$

Existing Techniques for Multi-Armed Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound
1. Stochastic $\leftrightarrow$ Adversarial (Bubeck and Slivkins, 2012)			✓	✓
2. EW + extra exploration (Slivkins and Seldin, 2014)			✓	✓
3. EW + adaptive learning rate (Ito et al., 2022)		✓		
4. FTRL + Tsallis entropy (Zimmert and Seldin, 2019; Ito, 2021)	✓	✓	✓	✓

Refined gap bound means obtaining $\sum_{i=1}^{K} \frac{\log T}{\Delta_i}$ instead of $\frac{K \log T}{\Delta_{\min}}$

Existing Techniques for Graph / Linear Bandits

Linear Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound
1. Stochastic ↔ Adversarial (Lee et al., 2021)				✓

Graph Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound
2. EW + extra exploration (Rouyer et al., 2022)			✓	✓
3. EW + adaptive learning rate (Ito et al., 2022)		✓		

Our Blackbox Approach

Standard FTRL ⟶

⟶ A best-of-three-world algorithm

adversarial: $\mathcal{O}(\sqrt{\beta T})$

stochastic: $\mathcal{O}\left(\frac{\beta \log T}{\Delta_{\min}}\right)$

corrupted: $\mathcal{O}\left(\frac{\beta \log T}{\Delta_{\min}} + \sqrt{\frac{\beta C \log T}{\Delta_{\min}}}\right)$

adversarial: $\mathcal{O}(\sqrt{\beta T})$

Assumption:
The best action/policy is *unique*
$\Delta_{\min} =$ gap between the best and the second-best action/policy

Improvement via Our Approach

Linear Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound
(Lee et al., 2021)				✓
Our Approach	✓	✓		

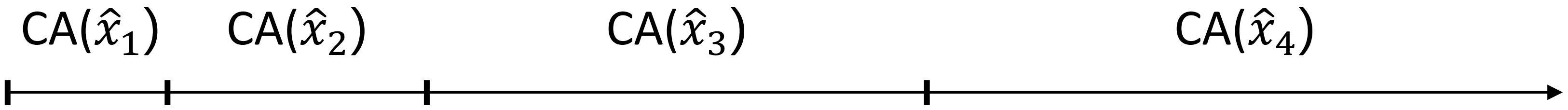
Graph Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal arms	Refined gap bound
(Rouyer et al., 2022)			✓	✓
(Ito et al., 2022)		✓		
Our Approach	✓	✓		

Contextual Bandits

	$\log T$	$\tilde{O}(\sqrt{C})$	Multiple optimal policies	Refined gap bound
Our Approach	✓	✓		

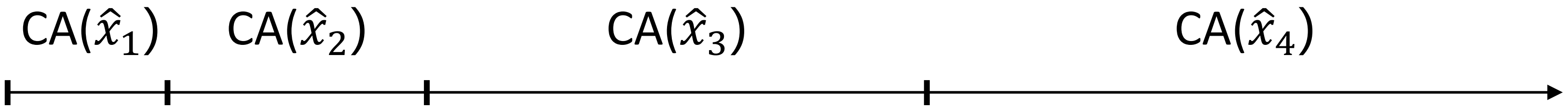
Our Approach



Each epoch runs a **candidate-aware** algorithm (CA) with candidate $\hat{x}_i \in \mathcal{X}$ as input.

action set

Our Approach



CA($\hat{x}_1$) CA($\hat{x}_2$) CA($\hat{x}_3$) CA($\hat{x}_4$)

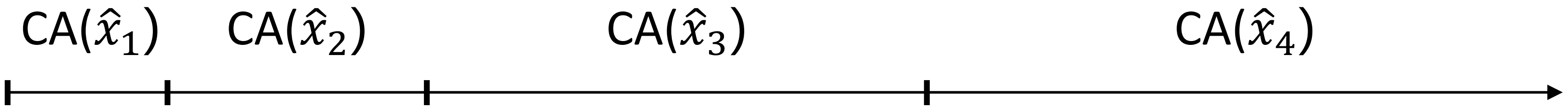
Each epoch runs a **candidate-aware** algorithm (CA) with candidate $\hat{x}_i \in \mathcal{X}$ as input.

action set

CA($\hat{x}$) need to

- Guarantee the standard $\sqrt{T}$ regret against all actions in $\mathcal{X}$
- Guarantee *an improved regret bound* against $\hat{x}$

Our Approach



CA($\hat{x}_1$) CA($\hat{x}_2$) CA($\hat{x}_3$) CA($\hat{x}_4$)

Each epoch runs a **candidate-aware** algorithm (CA) with candidate $\hat{x}_i \in \mathcal{X}$ as input.

action set

CA($\hat{x}$) need to

- Guarantee the standard $\sqrt{T}$ regret against all actions in $\mathcal{X}$
- Guarantee *an improved regret bound* against $\hat{x}$

Below, we will explain

1. The precise meaning of the *improved regret bound*, and the implementation of CA($\hat{x}$)
2. When to start a new epoch, and how to decide $\hat{x}_i$

1. The Requirement for CA($\hat{x}$)

Given an action $\hat{x}$ as input, CA($\hat{x}$) needs to ensure

$$\sum_{t=1}^{T}(\ell_t(x_t) - \ell_t(x)) \le \begin{cases} \sqrt{\beta T} & \text{For all } x \\ \sqrt{\beta \sum_{t=1}^{T}(1 - p_t(\hat{x}))\ \log T} & \text{if } x = \hat{x} \end{cases}$$

Probability of choosing $\hat{x}$ at round t

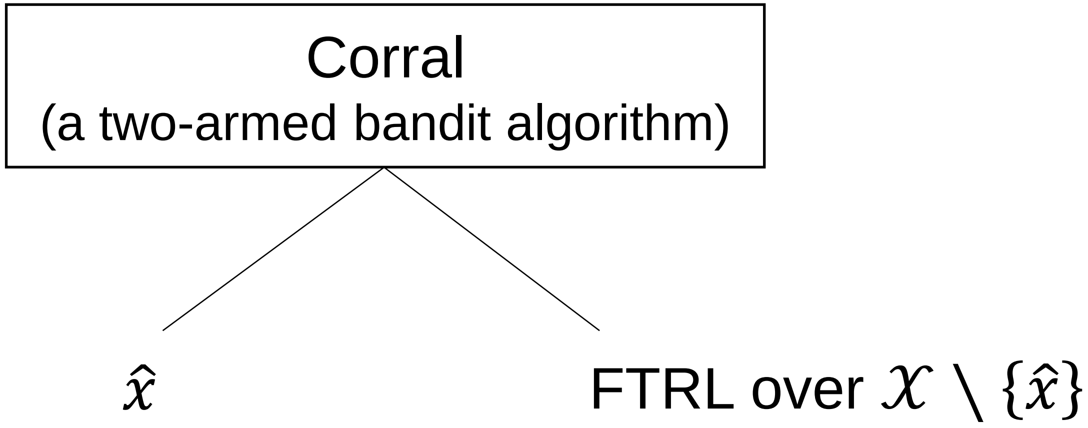
1. The Requirement for CA($\hat{x}$)

Given an action $\hat{x}$ as input, CA($\hat{x}$) needs to ensure

$$\sum_{t=1}^{T}\left(\ell_t(x_t)-\ell_t(x)\right) \le \begin{cases} \sqrt{\beta T} & \text{For all } x \\ \sqrt{\beta \sum_{t=1}^{T}\left(1-p_t(\hat{x})\right)\ \log T} & \text{if } x=\hat{x} \end{cases}$$

Probability of choosing $\hat{x}$ at round t

Implementation:



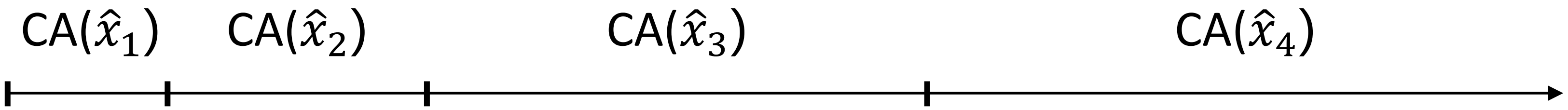
Corral

(a two-armed bandit algorithm)

$\hat{x}$

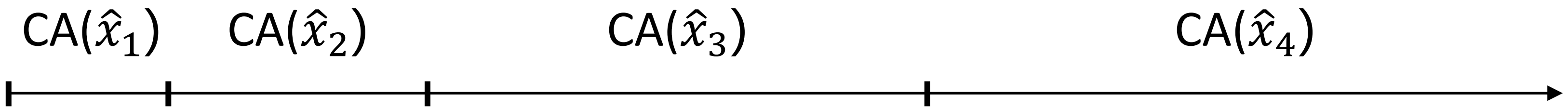
FTRL over $\mathcal{X} \setminus \{\hat{x}\}$

2. Epoch Scheduling and Candidate Assignment



Epoch i terminates if both of the following hold:

2. Epoch Scheduling and Candidate Assignment

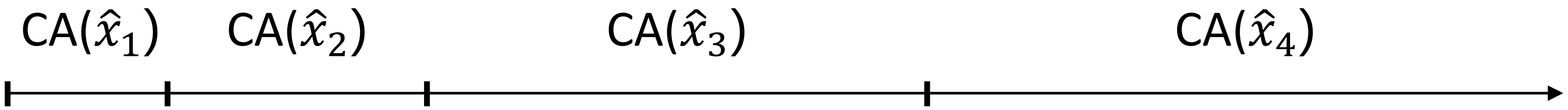


CA($\hat{x}_1$) CA($\hat{x}_2$) CA($\hat{x}_3$) CA($\hat{x}_4$)

Epoch i terminates if both of the following hold:

- There exists $x \neq \hat{x}_i$ chosen more than half of the times in epoch i (This x is then set as $\hat{x}_{i+1}$)

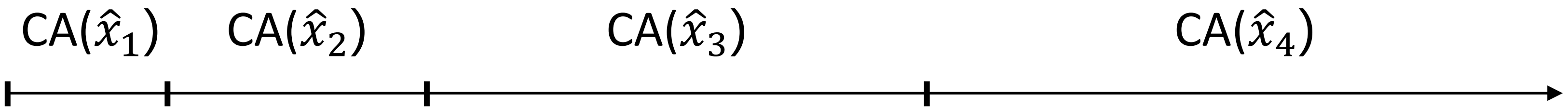
2. Epoch Scheduling and Candidate Assignment



Epoch i terminates if both of the following hold:

- There exists $x \neq \hat{x}_i$ chosen more than half of the times in epoch i (This x is then set as $\hat{x}_{i+1}$)
- Length(epoch i) > 2 × Length(epoch $i-1$) (for $i > 1$)

2. Epoch Scheduling and Candidate Assignment



CA($\hat{x}_1$) CA($\hat{x}_2$) CA($\hat{x}_3$) CA($\hat{x}_4$)

Epoch i terminates if both of the following hold:

- There exists $x \neq \hat{x}_i$ chosen more than half of the times in epoch i (This x is then set as $\hat{x}_{i+1}$)
- Length(epoch i) $>$ 2 × Length(epoch $i-1$) (for $i > 1$)

Theorem:

The overall procedure guarantees

$$\sum_{t=1}^{T} \left(\ell_t(x_t) - \ell_t(x)\right) \leq \begin{cases} \dfrac{\beta \log T}{\Delta_{\min}} + \sqrt{\dfrac{\beta \log T}{\Delta_{\min}} \cdot C} & \text{in the stochastic/corrupted world} \\ \sqrt{\beta T} & \text{in the adversarial world} \end{cases}$$

Summary

- We provide a general way to convert an **FTRL** to a **best-of-three-world** algorithm.

- The conversion achieves two of the four desired properties in a wide range of settings, producing state-of-the-art results in **graph / linear / contextual bandits.**

☑ $\sqrt{C \log T}$ ☑ $\log T$ ☒ Multiple optimal actions ☒ Refined gap bound

- Future work: handling multiple optimal actions and achieving refined gap bound