Learning on the Edge: Online Learning with Stochastic Feedback Graphs

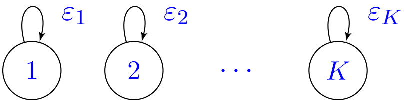
Emmanuel Esposito[1,2] Federico Fusco[3] Dirk van der Hoeven[1] Nicolò Cesa-Bianchi[1]

[1]Università degli Studi di Milano [2]Istituto Italiano di Tecnologia [3]Sapienza Università di Roma

A Concrete Example

Faulty bandits:

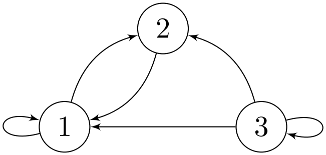
- Central agent repeatedly performing a decision-making task (e.g., daily)
- Sensors $s_1, \dots, s_K$ communicating daily with the agent
- Every day, agent sends a measurement request to some sensor s_i
- Communication with s_i fails independently w.p. $1 - \varepsilon_i$
- If the request is accepted, s_i sends back a measurement



Feedback Graph

Finite set of actions $V = \{1, \ldots, K\}$.

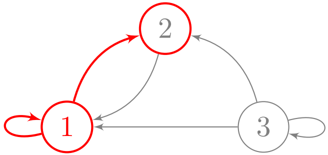
A directed graph $G = (V, E)$ over actions determines the feedback structure.



Feedback Graph

Finite set of actions $V = \{1, \ldots, K\}$.

A directed graph $G = (V, E)$ over actions determines the feedback structure.



At any time t, the choice $I_t \in V$ allows to observe actions in $N_G^{\mathrm{out}}(I_t) = \{i \in V : (I_t, i) \in E\}$.

Online Learning with Stochastic Feedback Graphs

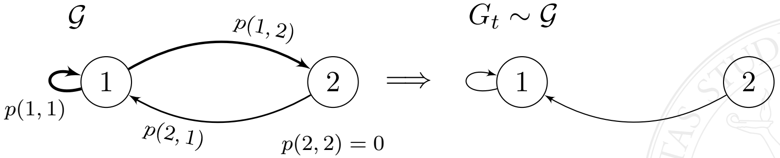
At each round $t = 1, \ldots, T$:

- learner plays action $I_t \sim \pi_t$

Online Learning with Stochastic Feedback Graphs

At each round $t = 1, \ldots, T$:

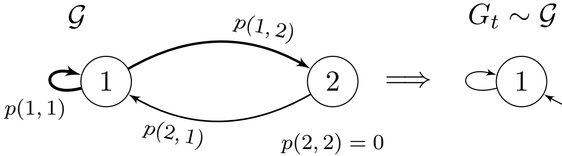
- learner plays action $I_t \sim \pi_t$
- environment generates $G_t = (V, E_t) \sim \mathcal{G}$



Online Learning with Stochastic Feedback Graphs

At each round $t = 1, \dots, T$:

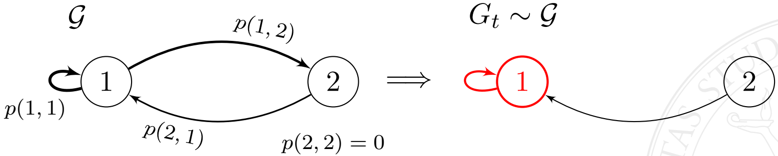
- learner plays action $I_t \sim \pi_t$
- environment generates $G_t = (V, E_t) \sim \mathcal{G}$
- learner incurs loss $\ell_t(I_t) \in [0, 1]$ and observes $\{\ell_t(i) : i \in N^{\text{out}}_{G_t}(I_t)\}$



Online Learning with Stochastic Feedback Graphs

At each round $t = 1, \ldots, T$:

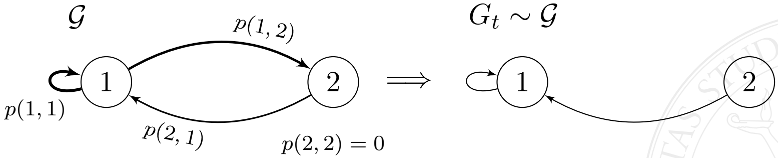
- learner plays action $I_t \sim \pi_t$
- environment generates $G_t = (V, E_t) \sim \mathcal{G}$
- learner incurs loss $\ell_t(I_t) \in [0, 1]$ and observes $\{\ell_t(i) : i \in N^{\text{out}}_{G_t}(I_t)\}$



Online Learning with Stochastic Feedback Graphs

At each round $t = 1, \ldots, T$:

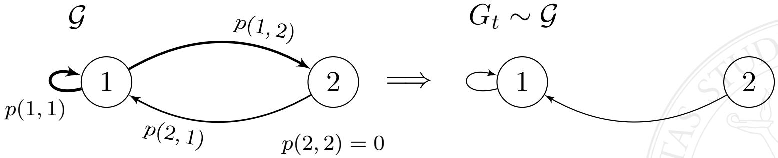
- learner plays action $I_t \sim \pi_t$
- environment generates $G_t = (V, E_t) \sim \mathcal{G}$
- learner incurs loss $\ell_t(I_t) \in [0, 1]$ and observes $\{\ell_t(i) : i \in N^{\text{out}}_{G_t}(I_t)\}$
- learner updates $\pi_t \mapsto \pi_{t+1}$



Online Learning with Stochastic Feedback Graphs

At each round $t = 1, \ldots, T$:

- learner plays action $I_t \sim \pi_t$
- environment generates $G_t = (V, E_t) \sim \mathcal{G}$
- learner incurs loss $\ell_t(I_t) \in [0, 1]$ and observes $\{\ell_t(i) : i \in N^{\text{out}}_{G_t}(I_t)\}$
- learner updates $\pi_t \mapsto \pi_{t+1}$

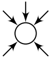


Goal: minimize regret

$$R_T = \max_{k \in V} \mathbb{E}\left[\sum_{t=1}^{T} (\ell_t(I_t) - \ell_t(k))\right]$$

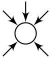
Online Learning with Feedback Graphs

Families of (deterministic) feedback graphs:

- **Strongly observable:** all actions  or 
  Regret: $\tilde{O}(\sqrt{\alpha T})$ where α is the independence number.

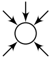
Online Learning with Feedback Graphs

Families of (deterministic) feedback graphs:

- **Strongly observable:** all actions  or 
  Regret: $\tilde{O}(\sqrt{\alpha T})$ where α is the independence number.
- **Weakly observable:** all actions observed but not strongly observable.
  Regret: $\tilde{O}(\delta^{1/3} T^{2/3})$ where δ is the weak domination number.

Online Learning with Feedback Graphs

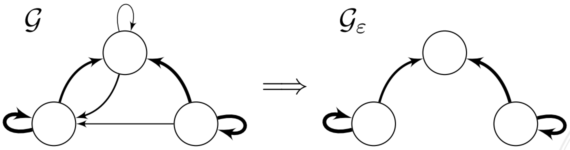
Families of (deterministic) feedback graphs:

- **Strongly observable:** all actions  or 
  Regret: $\tilde{O}(\sqrt{\alpha T})$ where α is the independence number.
- **Weakly observable:** all actions observed but not strongly observable.
  Regret: $\tilde{O}(\delta^{1/3}T^{2/3})$ where δ is the weak domination number.
- **Non-observable:** at least an action not observed.
  Regret: $\Omega(T)$.

Thresholding and Support

Consider a stochastic feedback graph $\mathcal{G} = \{p(i,j) : i,j \in V\}$.

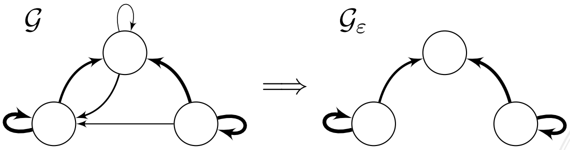
Thresholded stochastic feedback graph $\mathcal{G}_\varepsilon = \{p(i,j)\mathbb{I}_{\{p(i,j)\geq\varepsilon\}} : i,j \in V\}$



Thresholding and Support

Consider a stochastic feedback graph $\mathcal{G} = \{p(i,j) : i,j \in V\}$.

Thresholded stochastic feedback graph $\mathcal{G}_\varepsilon = \{p(i,j)\mathbb{I}_{\{p(i,j)\geq\varepsilon\}} : i,j \in V\}$



The support of $\mathcal{G}$ is $\mathrm{supp}(\mathcal{G}) = G = (V,E)$ where $E = \{(i,j) \in V \times V : p(i,j) > 0\}$.

Note: all "deterministic" (graph-theoretical) notions may extend to $\mathcal{G}$ via $\mathrm{supp}(\mathcal{G})$.

EdgeCatcher: From Stochastic to Deterministic

Let $\mathcal{A}$ be a learning algorithm for OL with deterministic feedback graph.

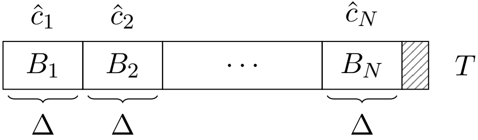
Initial round-robin to learn optimal threshold ε^* and a "good estimate" $\hat{\mathcal{G}}$ for $\mathcal{G}$.

EdgeCatcher: From Stochastic to Deterministic

Let $\mathcal{A}$ be a learning algorithm for OL with deterministic feedback graph.

Initial round-robin to learn optimal threshold ε^* and a "good estimate" $\hat{\mathcal{G}}$ for $\mathcal{G}$.

Blocks reduction: given threshold ε and $\hat{\mathcal{G}}$



Meta-instance for $\mathcal{A}$ with graph $\mathrm{supp}(\hat{\mathcal{G}}_\varepsilon)$ and losses $\hat{c}_1, \ldots, \hat{c}_N$.

EdgeCatcher: Regret Bound

The blocks reduction, given ε and $\hat{\mathcal{G}}$, achieves

$$R_T \le \Delta R_N^{\mathcal{A}}(\mathrm{supp}(\hat{\mathcal{G}}_\varepsilon)) + \Delta$$

Setting $\Delta = \Theta(\frac{1}{\varepsilon^*} \ln(KT))$, EdgeCatcher achieves

$$R_T = \tilde{O}\left(\min\left\{\min_{\varepsilon} \sqrt{(\alpha(\mathcal{G}_\varepsilon)/\varepsilon)T},\ \min_{\varepsilon}(\delta(\mathcal{G}_\varepsilon)/\varepsilon)^{1/3}T^{2/3}\right\}\right)$$

Nearly minimax-optimal in T, ε, and graph parameters.

$\textsc{otc}\mathrm{G}$: Be Optimistic If You Can, Commit If You Must

Assumption: observe G_t at the end of round t in addition to losses.

We design an algorithm based on $\textsc{Exp3.G}$ using new importance-weighted estimates $\tilde{\ell}_t(i)$ with upper confidence bounds $\hat{p}_t(j,i)$ for $p(j,i)$:

$$\tilde{\ell}_t(i) = \frac{\mathbb{I}_{\{I_t \to i \text{ in both } G_t \text{ and } \hat{G}_t\}}}{\sum_{j \xrightarrow{\hat{G}_t} i} \pi_t(j)\hat{p}_t(j,i)} \ell_t(i)$$

$\textsc{otcG}$: Be Optimistic If You Can, Commit If You Must

Assumption: observe G_t at the end of round t in addition to losses.

We design an algorithm based on $\textsc{Exp3.G}$ using new importance-weighted estimates $\tilde{\ell}_t(i)$ with upper confidence bounds $\hat{p}_t(j,i)$ for $p(j,i)$:

$$\tilde{\ell}_t(i) = \frac{\mathbb{I}_{\{I_t \to i \text{ in both } G_t \text{ and } \hat{G}_t\}}}{\sum_{j \xrightarrow{\hat{G}_t} i} \pi_t(j)\hat{p}_t(j,i)} \ell_t(i)$$

By an empirical Bernstein's bound,

$$\hat{p}_t(j,i) = \tilde{p}_t(j,i) + C_1 \sqrt{\frac{\ln(KT)}{t-1}\tilde{p}_t(j,i)} + C_2 \frac{\ln(KT)}{t-1}$$

where $\tilde{p}_t(j,i) = \frac{1}{t-1}\sum_{s=1}^{t-1} \mathbb{I}_{\{(j,i)\in E_s\}}$.

OTCG: Regret Bound

Optimistically assume strong observability, then commit to weak observability if better.

Regret:

$$R_T = \tilde{O}\left(\min\left\{\min_{\varepsilon}\sqrt{\alpha_{\mathsf{w}}(\mathcal{G}_\varepsilon)T},\ \min_{\varepsilon}\{\delta_{\mathsf{w}}(\mathcal{G}_\varepsilon)^{1/3}T^{2/3}+\sqrt{\sigma(\mathcal{G}_\varepsilon)T}\}\right\}\right)$$

α_{w} and δ_{w} are improved, weighted versions of α and δ containing the dependency on edge probabilities.

Conclusions and Future Work

- Our lower bounds show that EdgeCatcher and otcG are nearly minimax-optimal
- otcG improves with tighter graph-theoretical parameters

Conclusions and Future Work

- Our lower bounds show that EdgeCatcher and otcG are nearly minimax-optimal
- otcG improves with tighter graph-theoretical parameters

Questions:

- Can we use blocks of variable size in EdgeCatcher?
- Can we prove instance-dependent lower bounds matching the regret of otcG?
- Can we remove the need to observe the realized graph $G_t \sim \mathcal{G}$ in otcG?

Conclusions and Future Work

- Our lower bounds show that EdgeCatcher and otcG are nearly minimax-optimal
- otcG improves with tighter graph-theoretical parameters

Questions:

- Can we use blocks of variable size in EdgeCatcher?
- Can we prove instance-dependent lower bounds matching the regret of otcG?
- Can we remove the need to observe the realized graph $G_t \sim \mathcal{G}$ in otcG?

Thank you!