Nicolò Cesa-Bianchi
(University of Milan, Italy)

Online Learning with Feedback Graphs

Nicolò Cesa-Bianchi

Università degli Studi di Milano

Joint work with:

Noga Alon (Tel-Aviv University)
Ofer Dekel (Microsoft Research)
Tomer Koren (Technion and Microsoft Research)

Theory of repeated games

James Hannan
(1922–2010)

David Blackwell
(1919–2010)

Learning to play a game (1956)

Play a game repeatedly against a possibly suboptimal opponent

Prediction with expert advice

N actions

For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0, 1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)

Prediction with expert advice

N actions

For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0,1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)
2. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$

Prediction with expert advice

N actions

For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0,1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)
2. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$
3. Player gets feedback information: $\boldsymbol{\ell}_t = \big(\ell_t(1), \ldots, \ell_t(N)\big)$

Regret

The loss process $\langle \ell_t \rangle_{t \geqslant 1}$ is nonstochastic and unknown to the (randomized) player $I_1, I_2, \ldots$

Regret minimization

$$R_T \stackrel{\text{def}}{=} \mathbb{E}\left[\sum_{t=1}^{T} \ell_t(I_t)\right] - \min_{i=1,\ldots,N} \sum_{t=1}^{T} \ell_t(i) \stackrel{\text{want}}{=} o(T)$$

Stochastic asymptotic lower bound for experts' game

$$R_T = \big(1 - o(1)\big) \sqrt{\frac{T \ln N}{2}}$$

Exponentially weighted forecaster

At time t pick action $I_t = i$ with probability proportional to

$$\exp\left(-\eta \sum_{s=1}^{t-1} \ell_s(i)\right)$$

the sum at the exponent is the total loss of action i up to now

Regret bound [How to use expert advice, 1997]

If $\eta = \sqrt{\frac{\ln N}{8T}}$ then $R_T \leqslant \sqrt{\frac{T \ln N}{2}}$

Matching asymptotic lower bound including constants

The bandit problem: playing an unknown game

N actions

For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0, 1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)

The bandit problem: playing an unknown game

N actions

For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0, 1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)
2. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$

The bandit problem: playing an unknown game

N actions

For $t = 1, 2, \dots$

1. Loss $\ell_t(i) \in [0, 1]$ is assigned to every action $i = 1, \dots, N$ (hidden from the player)
2. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$
3. Player gets feedback information: Only $\ell_t(I_t)$ is revealed

The bandit problem: playing an unknown game

N actions

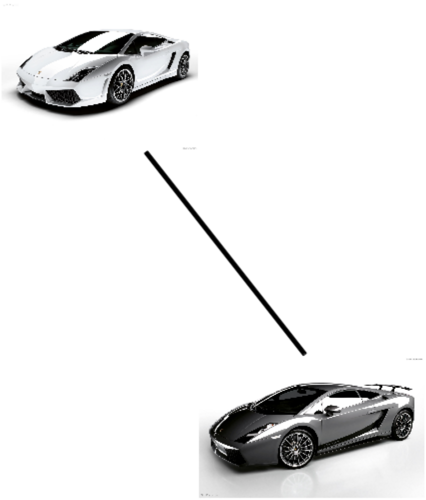
For $t = 1, 2, \ldots$

1. Loss $\ell_t(i) \in [0, 1]$ is assigned to every action $i = 1, \ldots, N$ (hidden from the player)
2. Player picks an action I_t (possibly using randomization) and incurs loss $\ell_t(I_t)$
3. Player gets feedback information: Only $\ell_t(I_t)$ is revealed

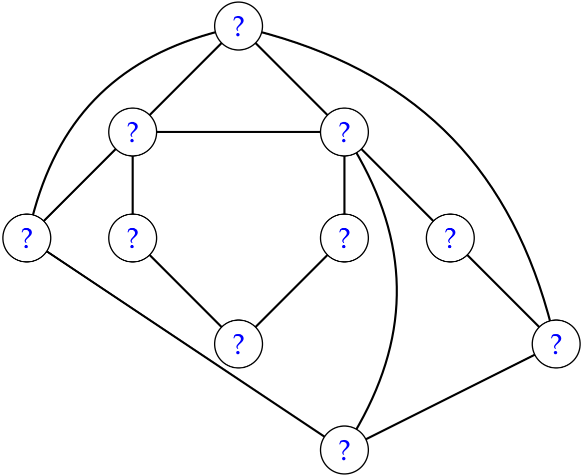
Many applications

Ad placement, recommender systems, online auctions, . . .

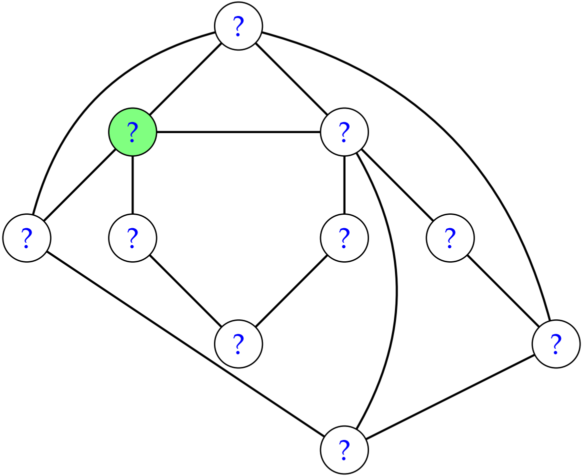
Relationships between actions



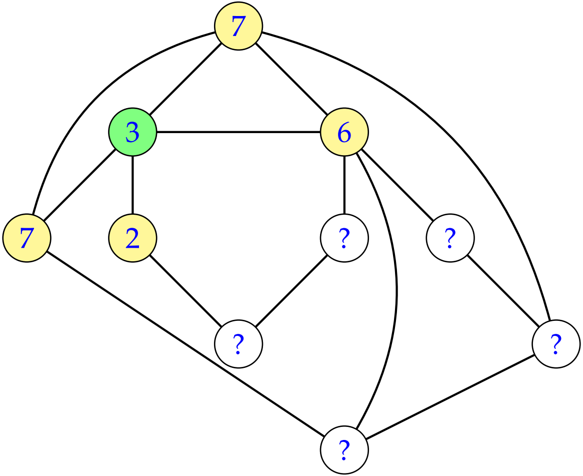
A graph of relationships over actions



A graph of relationships over actions

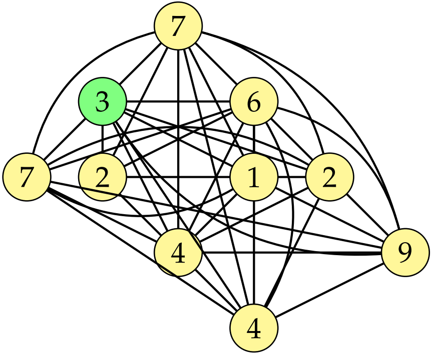


A graph of relationships over actions

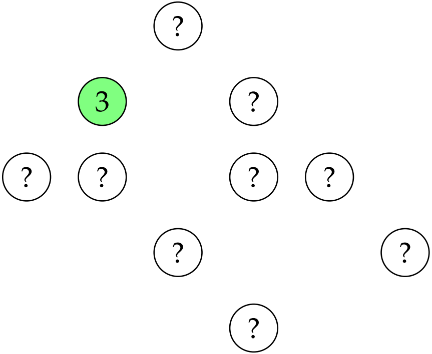


Recovering expert and bandit settings

Experts: clique



Bandits: empty graph



Exponentially weighted forecaster — Reprise

Player's strategy

- $\mathbb{P}_t(I_t = i) \quad \propto \quad \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) \qquad i = 1, \dots, N$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ observed}\big)} & \text{if } \ell_t(i) \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$

Importance sampling estimator

$\mathbb{E}_t\left[\widehat{\ell}_t(i)\right] = \ell_t(i)$ unbiasedness

$\mathbb{E}_t\left[\widehat{\ell}_t(i)^2\right] \leqslant \dfrac{1}{\mathbb{P}_t\big(\ell_t(i) \text{ observed}\big)}$ variance control

Regret bounds

Analysis (undirected graphs)

$$R_T \leqslant \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\mathbb{P}_t(i \text{ is played})}{\mathbb{P}_t(\ell_t(i) \text{ is observed})}$$

Lemma

For any undirected graph $G = (V, E)$ and for any probability assignment $p_1, \ldots, p_N$ over its vertices

$$\sum_{i=1}^{N} \frac{p_i}{p_i + \sum_{j \in N_G(i)} p_j} \leqslant \alpha(G)$$

$\alpha(G)$ is the independence number of G (largest subset of V such that no two distinct vertices in it are adjacent in G)

Regret bounds

Analysis (undirected graphs)

$$R_T \leqslant \frac{\ln N}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \alpha(G) = \sqrt{T\alpha(G) \ln N} \qquad \text{by choosing } \eta$$

Special cases

Experts (clique): $\alpha(G) = 1$ $\quad R_T \leqslant \sqrt{T \ln N}$

Bandits (empty graph): $\alpha(G) = N$ $\quad R_T \leqslant \sqrt{TN \ln N}$

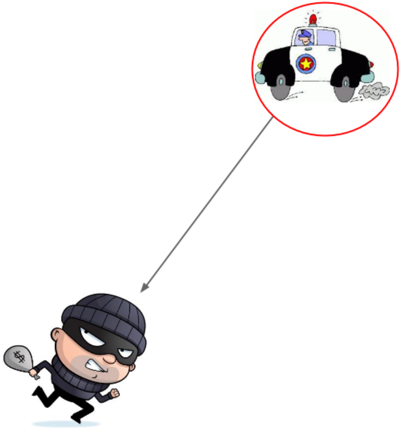
Minimax rate

The general bound is tight: $R_T = \Theta\left(\sqrt{T\alpha(G) \ln N}\right)$

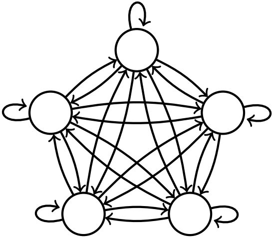
More general feedback models

Directed

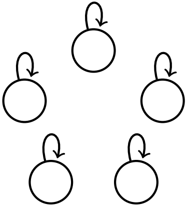
Interventions



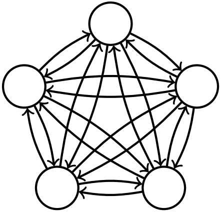
Old and new examples



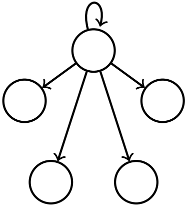
Experts



Bandits



Cops & Robbers



Revealing Action

Exponentially weighted forecaster with exploration

Player's strategy

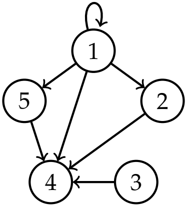
- $\mathbb{P}_t(I_t = i) \propto \frac{1-\gamma}{Z_t} \exp\left(-\eta \sum_{s=1}^{t-1} \widehat{\ell}_s(i)\right) + \gamma U_G \qquad i = 1, \dots, N$
- $\widehat{\ell}_t(i) = \begin{cases} \dfrac{\ell_t(i)}{\mathbb{P}_t\big(\ell_t(i) \text{ observed}\big)} & \text{if } \ell_t(i) \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$

U_G is uniform distribution supported on a subset of V

A characterization of feedback graphs

A vertex of G is:

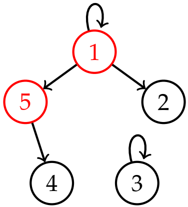
- observable if it has at least one incoming edge (possibly a self-loop)
- strongly observable if it has either a self-loop or incoming edges from all other vertices
- weakly observable if it is observable but not strongly observable



- 3 is not observable
- 2 and 5 are weakly observable
- 1 and 4 are strongly observable

Characterization of minimax rates

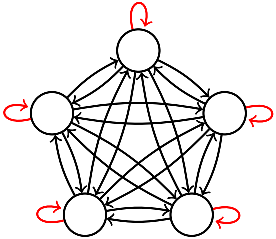
G is strongly observable	$R_T = \widetilde{\Theta}\left(\sqrt{\alpha(G)T}\right)$ u_G is uniform on V
G is weakly observable	$R_T = \widetilde{\Theta}\left(T^{2/3}\delta(G)\right)$ u_G is uniform on a weakly dominating set
G is not observable	$R_T = \Theta(T)$



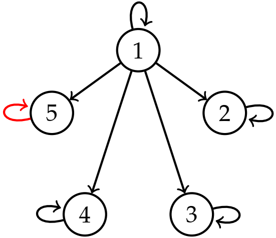
Weakly dominating set

$\delta(G)$ is the size of the smallest set that dominates all weakly observable nodes of G

Some curious cases



Presence of red loops does not affect minimax regret
$R_T = \Theta\left(\sqrt{T \ln N}\right)$



With red loop: strongly observable with $\alpha(G) = N - 1$
$R_T = \widetilde{\Theta}\left(\sqrt{NT}\right)$

Without red loop: weakly observable with $\delta(G) = 1$
$R_T = \widetilde{\Theta}\left(T^{2/3}\right)$

Conclusions

- An abstract, graph-theoretic framework for studying learning with partial feedback
- Sharp characterizations of rates in terms of graph-theoretic quantities
- Strong connections to the partial monitoring framework in game theory