A near-optimal algorithm for finite partial-monitoring games against adversarial opponents

Gábor Bartók
ETH Zürich

COLT 2013
Partial-Monitoring Games

learner environment
referee
action \( I_t \) outcome \( J_t \)
loss \( \ell_t = L(I_t, J_t) \)
feedback \( h_t = H(I_t, J_t) \)

Gábor Bartók (ETH Zürich)
Measure of Performance of the Learner

- How fast does the average cost converge to the optimal average cost?
- Compare the total losses $\Rightarrow$ **Regret**:

$$R_T = \sum_{t=1}^{T} L(I_t, J_t) - \inf_{i \in \mathcal{X}} \sum_{t=1}^{T} L(i, J_t).$$
Recent previous work

Classification of finite problems (Bartók et al., 2011)

Against iid opponents, the minimax regret of any finite partial-monitoring problem scales as either \( \{0, \frac{T^{1/2}}{T}, \frac{T^{2/3}}{T}, T\} \).

Same classification holds for adversarial environments (Foster and Rakhlin, 2012)

Foster and Rakhlin (2012): NeighborhoodWatch
Two-level Exponential Weights algorithm
- Chooses a local game randomly
- Chooses an action randomly

\( \tilde{O}(N\sqrt{T}) \) regret
New algorithm **GloGalExp3**

Chooses point-local game **deterministically**
Plays point-local game randomly (**Exp3**)
Construct set of neighboring action pairs $\mathcal{M}$
Explore them
Use difference estimates to constrain $p$ in all directions
Play the local game that “surely” contains $p$
Choice of the local game: not random
New algorithm II: Choosing the local game

Construct set of neighboring action pairs $\mathcal{M}$
Explore them
Use difference estimates to constrain $p$ in all directions
Play the local game that “surely” contains $p$

Choice of the local game: not random

Cell decomposition of the probability simplex; cell of action: region of optimality
New algorithm II: Choosing the local game

Construct set of neighboring action pairs $\mathcal{M}$
Explore them
Use difference estimates to constrain $p$
in all directions
Play the local game that “surely” contains $p$
Choice of the local game: not random

Cell decomposition of the probability simplex; cell of action: region of optimality
New algorithm II: Choosing the local game

Construct set of neighboring action pairs $\mathcal{M}$
Explore them
Use difference estimates to constrain $p$
in all directions
Play the local game that “surely” contains $p$
Choice of the local game: not random

Cell decomposition of the probability simplex; cell of action: region of optimality
New algorithm III: Playing the local game

Point-local games

Update method basically same as that of Foster and Rakhlin (2012)
- Draw actions $I$ and $I'$ from same distribution
- Choose action $I$, compare it to $I'$
- Loss estimate for update: $\hat{\ell}_i \leftarrow \left( \frac{\mathbb{I}\{I=i\}}{p_i} v_{i,I'} - v_{I,i} \right)^\top g$
- $E[\hat{\ell}_i - \hat{\ell}_j] = l_i - l_j$
Main theorem

Theorem

Given a locally observable finite partial monitoring game \( G = (L, H) \), algorithm \texttt{GlobalExp3} with appropriately set parameters achieves expected regret

\[
E[R_T] \leq 1 + 24 \frac{V_{\text{max}}}{\varepsilon_G} \sqrt{|\mathcal{M}| T \log(2T^2/|\mathcal{M}|)} + \sqrt{\frac{6}{\varepsilon_G}(2L_{\text{max}} + 4V_{\text{max}}) \sqrt{N' T \log N' \log T}}.
\]

\[
\tilde{O}\left(\frac{1}{\varepsilon_G} \sqrt{(N' + |\mathcal{M}|) T}\right)
\]

\(|\mathcal{M}| \leq \min(M - 1, N)\)
Thank you!
