Algorithmic chaining and the role of partial feedback in online nonparametric learning

Nicolò Cesa-Bianchi [1], Pierre Gaillard [2], Claudio Gentile [3] and Sébastien Gerchinovitz [4]

[1]Università degli Studi di Milano, Milano, Italy

[2]INRIA - Sierra Project-team, École Normale Supérieure, Paris, France

[3]Università degli Studi dell'Insubria, Varese, Italy

[4]Université Toulouse III - Paul Sabatier, Toulouse, France

Setting: online nonparametric (contextual) learning

Online protocol: at each round $t \in \mathbb{N}^*$,

1. The environment reveals a context vector $x_t \in [0,1]^d$.
2. The learner chooses an action $\widehat{y}_t \in [0,1]$.
3. The learner suffers the loss $\ell_t(\widehat{y}_t)$ and obtains feedback:
   - the instantaneous loss $\ell_t(\widehat{y}_t)$ in the bandit setting
   - $\ell_t(y)$ for all $y \geqslant \widehat{y}_t$ in the one-sided feedback setting
   - the loss function ℓ_t in the full-information setting.

Goal: denoting by $\mathcal{F}$ the set of 1-Lipschitz functions from $[0,1]^d$ to $[0,1]$ (for some given norms), we want to minimize the regret

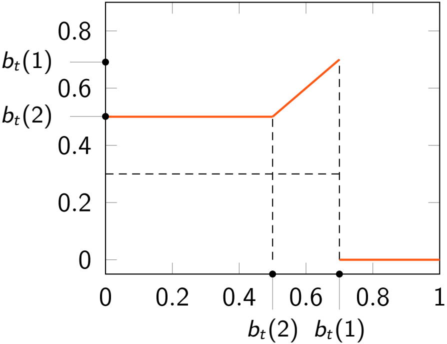
$$\mathrm{Reg}_T(\mathcal{F}) \ := \ \sum_{t=1}^{T} \ell_t\big(\widehat{y}_t\big) - \inf_{f \in \mathcal{F}} \sum_{t=1}^{T} \ell_t\big(f(x_t)\big)\,.$$

The sequence $(x_t, \ell_t)_{t \geqslant 1}$ is arbitrary and fixed at the beginning of the game; the loss functions ℓ_t satisfy a Lipschitz-type assumption.

An example of one-sided full information feedback

Online second-price auctions with reserve price:

- Seller with hidden reserve price $\widehat{y}_t$ (minimal revenue), multiple bidders
- The highest bidder wins the auction but pays the minimum of the second-highest bid and the reserve price (unless the reserve price is too big and the auction is lost).



If the seller observes the highest bid along with her revenue, she can compute $\ell_t(y)$ for all $y \geqslant \widehat{y}_t \rightsquigarrow$ one-sided full information feedback.

Full information feedback (suboptimal approach)

The learner observes $\ell_t(\cdot)$ at the end of round t.

Natural but suboptimal approach: discretize the set $\mathcal{F}$ of 1-Lipschitz functions from $[0,1]^d$ to $[0,1]$:

- in sup norm, this set can be approximated at precision ε with roughly $N_\varepsilon \approx 2^{(1/\varepsilon)^d}$ functions;
- using a classical mixture algorithm (like Hedge) on these N_ε experts yields regret

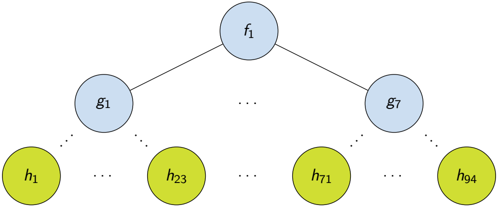
$$\begin{aligned}\mathrm{Reg}_T(\mathcal{F}) &\lesssim \sqrt{T \ln N_\varepsilon} + T\varepsilon && \text{if the } \ell_t \text{ are (semi-)Lipschitz} \\ &\lesssim T^{(d+1)/(d+2)} && \text{when optimizing in } \varepsilon\end{aligned}$$

Why suboptimal? We treat the functions in the discretization as uncorrelated experts, which is too pessimistic and harmful when $\mathcal{F}$ is large.

Full information: less pessimistic approach via chaining

Build a hierarchy of discretizations:

- the level-m discretization approximates $\mathcal{F}$ with precision 2^{-m};
- each level-m node is connected to its closest level-$(m-1)$ node;



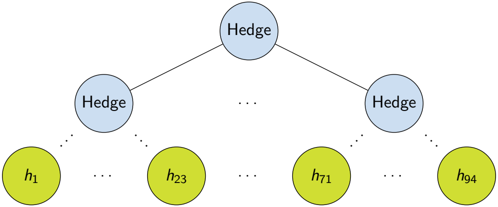
Hierarchical Hedge algorithm:

- each leaf h recommends its own (discretized) action $h(x_t)$;
- each internal node hosts an instance of Hedge using its children as experts; its regret is at most of order $2^{-m}\sqrt{T \ln N_{2^{-m}}}$ at level m since its children's losses are 2^{-m}-close.

Full information: less pessimistic approach via chaining

Build a hierarchy of discretizations:

- the level-m discretization approximates $\mathcal{F}$ with precision 2^{-m};
- each level-m node is connected to its closest level-$(m-1)$ node;



Hierarchical Hedge algorithm:

- each leaf h recommends its own (discretized) action $h(x_t)$;
- each internal node hosts an instance of Hedge using its children as experts; its regret is at most of order $2^{-m}\sqrt{T \ln N_{2^{-m}}}$ at level m since its children's losses are 2^{-m}-close.

Chaining (continued)

Summing the local regret bounds over any path in the tree, we obtain a regret bound of

$$\mathrm{Reg}_T(\mathcal{F}) \lesssim \sum_{m=0}^{M-1} 2^{-m}\sqrt{T \ln N_{2^{-m}}} + 2^{-M}T \approx \begin{cases} \sqrt{T} & \text{if } d \in \{1,2\} \\ T^{(d-1)/d} & \text{if } d \geqslant 3. \end{cases}$$

Remarks:

- Same upper bound as the one proven by Rakhlin et al. (2015) in a nonconstructive manner.
- Matches the lower bound of Hazan and Megiddo (2007).
- Our approach generalizes that of Cesa-Bianchi and Lugosi (1999) and Gaillard and Gerchinovitz (2015) to nonconvex Lipschitz losses.
- A variant of this hierarchical composition of Hedge instances can be implemented in polynomial time.

Can we use chaining for other feedbacks?

Bandit feedback

- Bad news: deriving regret bounds that scale as the effective range of the arms' losses, which was key for full information, is not possible in general for adversarial bandits (Gerchinovitz and Lattimore, 2016).
- Regret bounds : $T^{(d+2)/(d+3)}$ for semi-Lipschitz losses or $T^{(d+1)/(d+2)}$ for convex Lipschitz losses. See also the work of Slivkins (2014).

One-sided full-information feedback

- This stronger feedback, together with Lipschitzness of the losses, enables us to derive a regret bound for a variant of Exp4 that scales as the effective range of the arms' losses.
- Hierarchical algorithm: in the earlier tree, we replace Hedge with Exp4. We obtain a regret of order $T^{d/(d+1)}$ or even $T^{(d-1/3)/(d+2/3)}$ with an additional hierarchical penalization trick.

A brief summary of our results

More details at the poster session!

Feedback model	**Loss functions**	**Upper bound**
Bandit	Lipschitz	$T^{\frac{d+2}{d+3}}$
	Convex	$T^{\frac{d+1}{d+2}}$
One-sided full information	Semi-Lipschitz	$T^{\frac{d+1}{d+2}}$
	Lipschitz	$T^{\frac{d-1/3}{d+2/3}}$
Full information	Lipschitz	$T^{\frac{d-1}{d}}$

Table: Some regret bounds obtained in this paper (up to log factors).

References

Nicolò Cesa-Bianchi and Gábor Lugosi. On prediction of individual sequences. *The Annals of Statistics*, 27(6):1865–1895, 1999.

Pierre Gaillard and Sebastien Gerchinovitz. A chaining algorithm for online nonparametric regression. In *Proceedings of COLT'15*, volume 40, pages 764–796. JMLR: Workshop and Conference Proceedings, 2015.

Sébastien Gerchinovitz and Tor Lattimore. Refined lower bounds for adversarial bandits. In *Advances in Neural Information Processing Systems 29 (NIPS'16)*, pages 1198–1206, 2016.

Elad Hazan and Nimrod Megiddo. Online learning with prior knowledge. In *International Conference on Computational Learning Theory (COLT'07)*, pages 499–513. 2007.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *Journal of Machine Learning Research*, 16:155–186, 2015.

Aleksandrs Slivkins. Contextual bandits with similarity information. *Journal of Machine Learning Research*, 15(1):2533–2568, 2014.