Master Algorithms for Active Experts Problems based on Increasing Loss Values

or

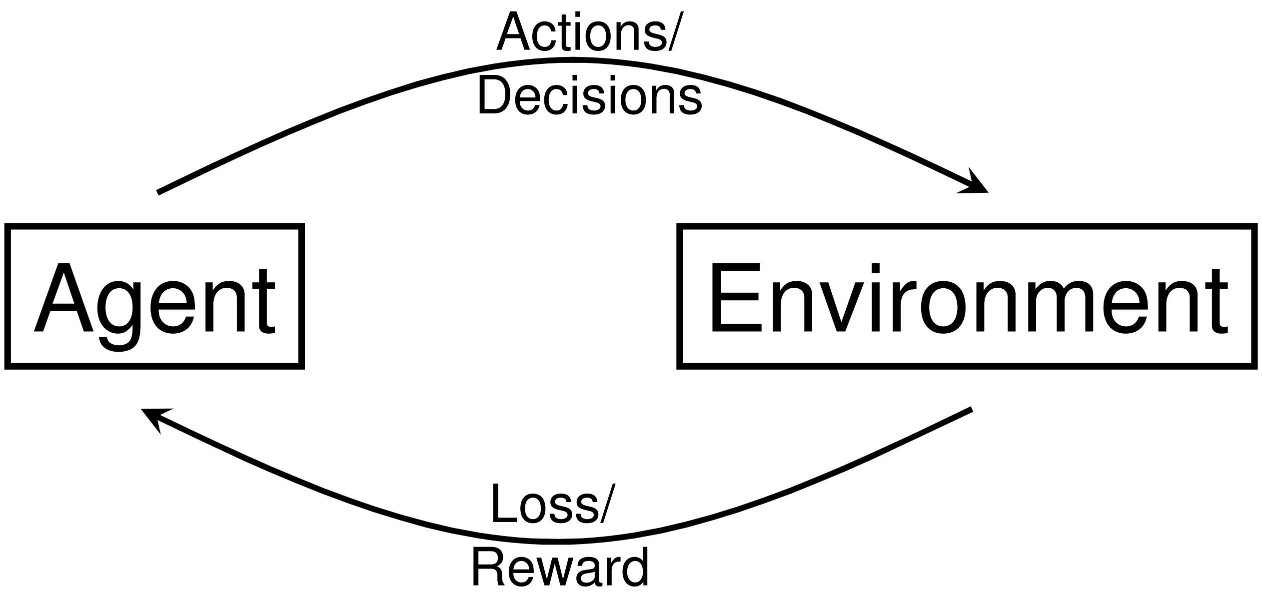
Defensive Universal Learning

Jan Poland and Marcus Hutter

IDSIA • Lugano • Switzerland

What

Universal learning: Algorithm that performs asymptotically optimal for any online decision problem.



How

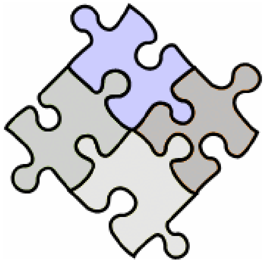
Prediction with
Expert Advice

Growing
Losses

Bandit
techniques

Infinite Expert classes

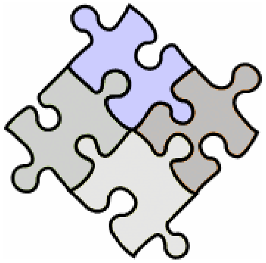
Prediction with Expert Advice



t	=	1	2	3	4	5 ...
Expert 1: *loss*	=	1	0	1	0	
Expert 2: *loss*	=	½	1	1	0	
Expert 3: *loss*	=	0	0	½	1	
...						
Master: *loss*	=	1	0	½	0	

instantaneous losses are bounded in [0,1]

Prediction with Expert Advice



Do not follow the leader:

$$\begin{array}{lcccccc} \text{Expert 1: } loss & = & 0 & 1 & 0 & 1 & \\ & & & & & & \ldots \\ \text{Expert 2: } loss & = & \tfrac{1}{2} & 0 & 1 & 0 & \end{array}$$

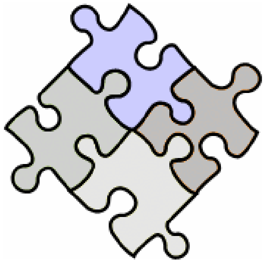
... but the perturbed leader:

$$\begin{aligned} &Regret \\ \Rightarrow \quad &= \mathbf{E}\, loss(Master) - loss(Best\ expert) \\ &= O(\sqrt{t}\log n) \end{aligned}$$

proven for adversarial environments!

[Hannan, Littlestone, Warmuth, Cesa-Bianchi, McMahan, Blum, etc.]

Learning Rate



Cumulative loss grows

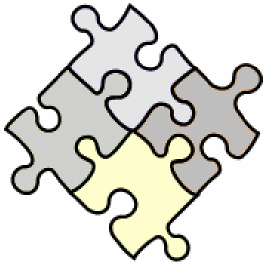
$\Rightarrow$ has to be scaled down
(otherwise we end up following the leader)

$\Rightarrow$ dynamic learning rate $1/\sqrt{t}$

learning rate and bounds can be significantly improved for small losses
$\Rightarrow$ things get more complicated

[Cesa-Bianchi et al., Auer et al., etc.]

Priors for Expert Classes



Expert class of finite size n

$\Rightarrow$ uniform prior $\equiv 1/n$ is common

$\Rightarrow$ uniform complexity $\equiv \log n$

Countably infinite expert class $n = \infty$

$\Rightarrow$ uniform prior impossible

$\Rightarrow$ bounds are instead in $\log w_i^{-1}$

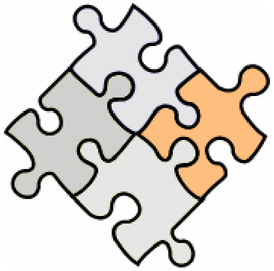
i is the best expert in hindsight

[Hutter, Poland for dynamic learning rate]

Universal Expert Class

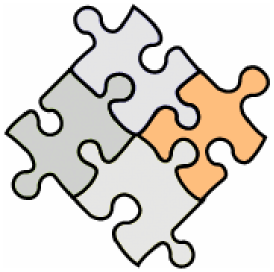
- Expert i = ith program on some universal Turing machine
- Prior complexity = length of the program
- Interprete the output appropriately, depending on the problem
- This construction is common in Algorithmic Information Theory

Bandit Case



		t =	1	2	3	4	5 ...
Expert 1:	*loss*	=	1	?	?	?	
Expert 2:	*loss*	=	?	?	?	0	
Expert 3:	*loss*	=	?	0	½	?	
...							
Master:	*loss*	=	1	0	½	0	

Bandit Case



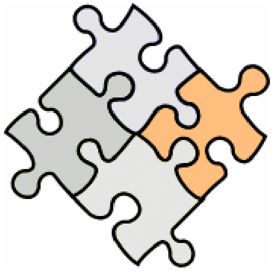
- Explore with small probability γ_t , otherwise exploit
- γ_t is the exploration rate
- $\gamma_t \to 0$ as $t \to \infty$
- Deal with estimated losses:

$$\frac{\textit{observed loss of the action}}{\textit{probability of the action}}$$

- unbiased estimate

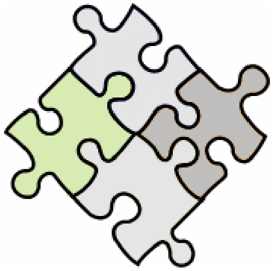
[Auer et al., McMahan and Blum]

Bandit Case: Consequences



- Bounds are in w_i instead of $-\log w_i$
- this (exponentially worse) bound is sharp in general
- Analysis gets harder for adaptive adversaries (with martingales...)

Reactive Environments



Repeated game: Prisoner‘s Dilemma

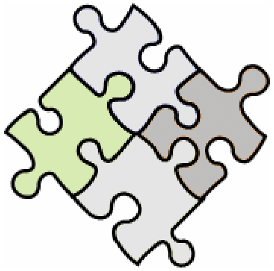
	C	D
Cooperate (C)	0.3	1
Defect (D)	0	0.7

$$\text{Cooperate:} \quad loss \; = \; \boxed{0.3} \quad 0.3 \quad 1 \quad \boxed{1} \quad \boxed{0.3}$$

$$\text{Defect:} \quad loss \; = \; 0 \quad \boxed{0} \quad \boxed{0.7} \quad 0.7 \quad 0$$

[de Farias and Megiddo, 2003]

Prisoner's Dilemma



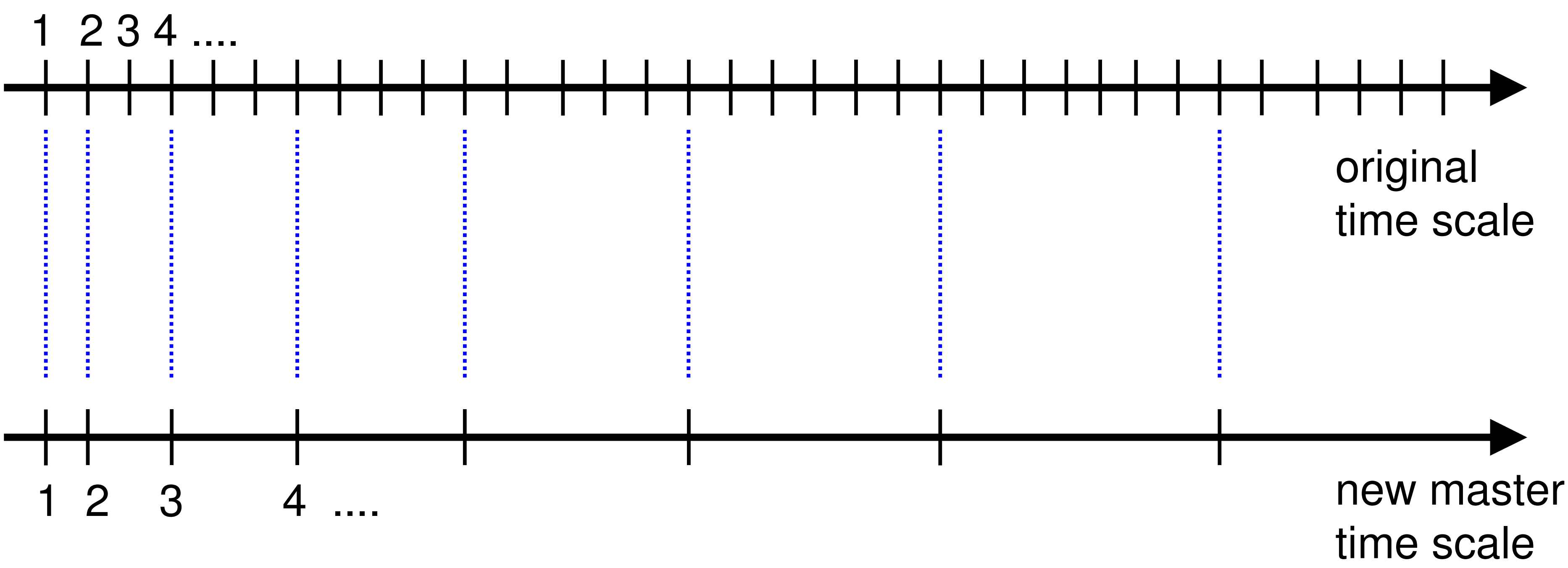
- defecting is dominant
- but still cooperating may have the better long term reward
- e.g. against "Tit for Tat"
- Expert Advice fails against Tit for Tat
- Tit for Tat is reactive

Cooperate: $loss$	=	0.3	0.3	1	1	0.3
Defect: $loss$	=	0	0	0.7	0.7	0

[de Farias and Megiddo, 2003]

Time Scale Change

Idea: Yield control to selected expert for increasingly many time steps



⇒ instantaneous losses may grow in time

Follow or Explore (FoE)

Need master algorithm +analysis for

- losses in $[0,B_t]$, B_t grows
- countable expert classes
- dynamic learning rate
- dynamic exploration rate
- technical issue: dynamic confidence for almost sure assertions

$\Rightarrow$ Algorithm FoE (Follow or Explore)

(details are in the paper)

Main Result

Theorem: For any online decision problem, FoE performs in the limit as well as any computable strategy (expert). That is, FoE‘s average per round regret converges to 0.

Moreover, FoE uses only finitely many experts at each time, thus is computationally feasible.

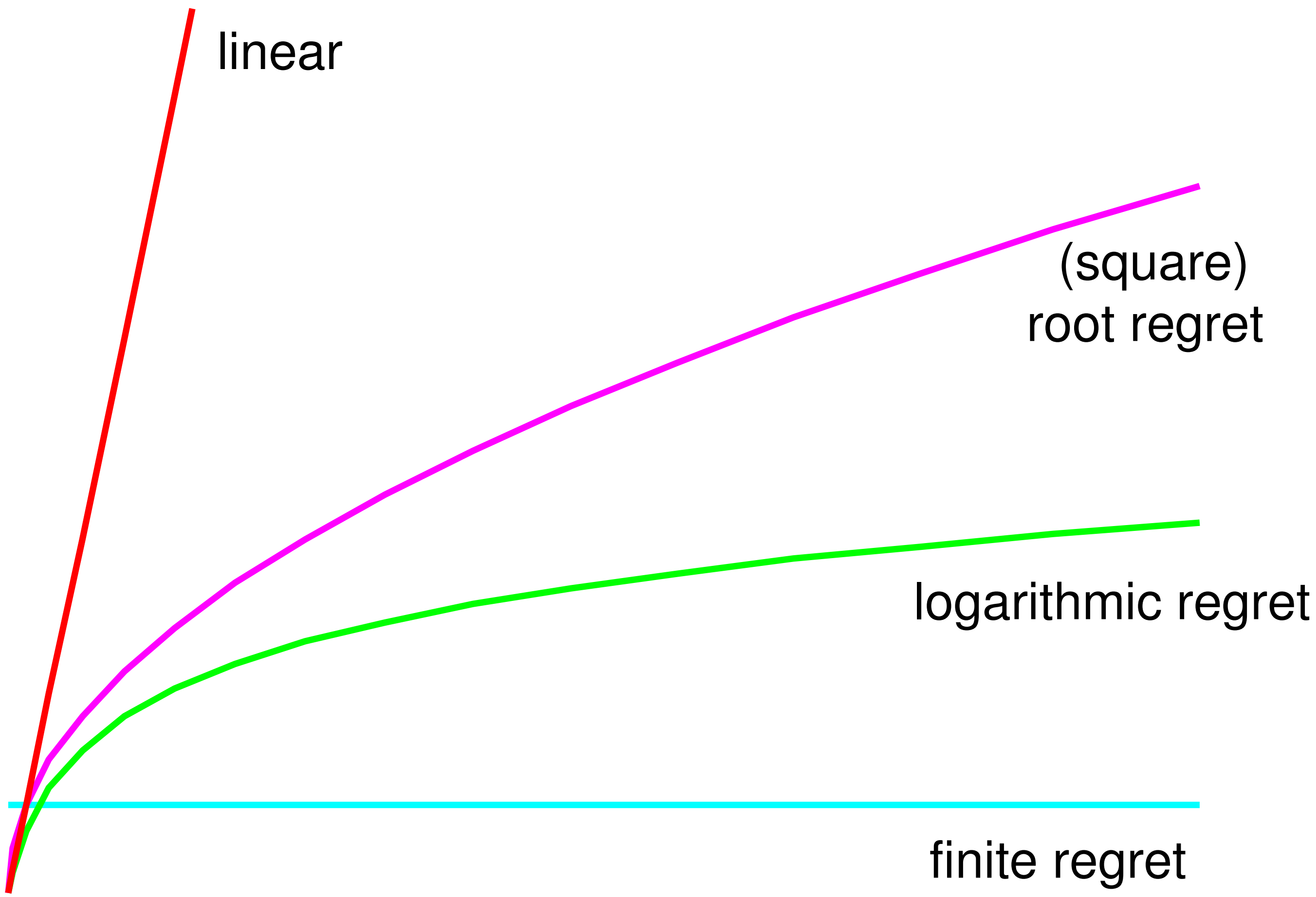
Universal Optimality

universal optimal = the average per-round regret tends to zero

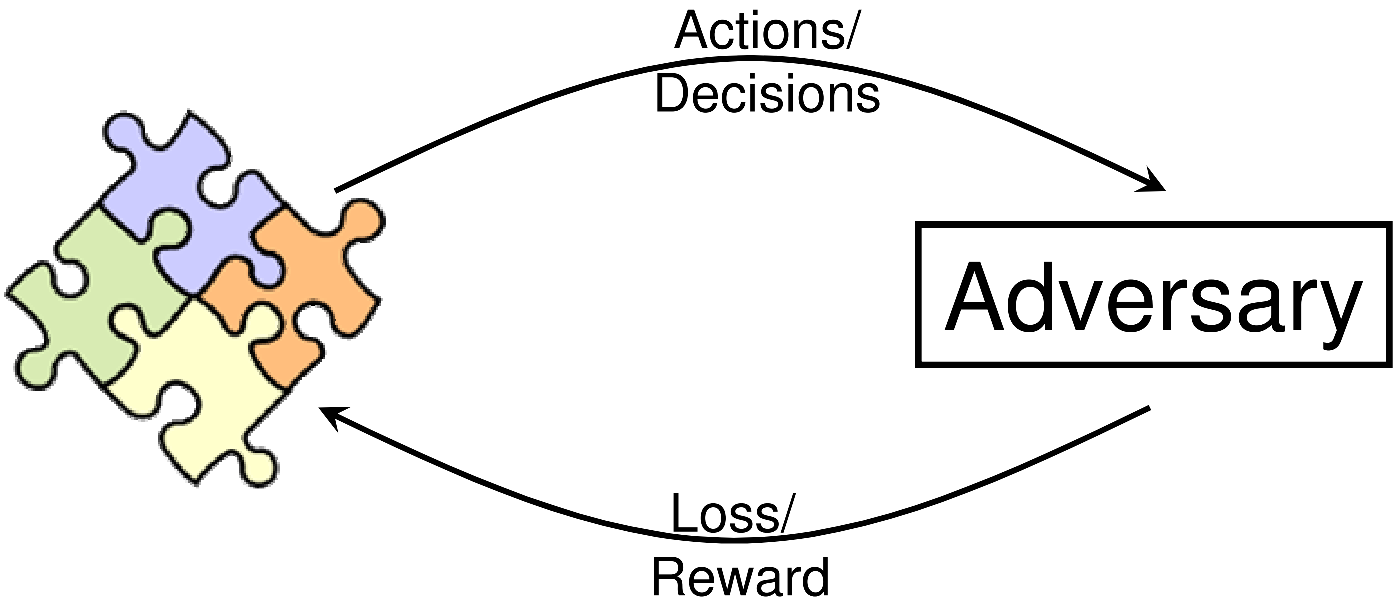
$\Leftrightarrow$

the cumulative regret grows slower than t

Universal Optimality



Conclusions



- FoE is universally optimal in the limit
- but maybe too defensive for practice?!

Thank you!