Competing with stationary prediction strategies

Vladimir Vovk

Department of Computer Science
Royal Holloway, University of London
Egham, Surrey, England

vovk@cs.rhul.ac.uk

San Diego, June 15, 2007

Game of prediction

Environment announces $(\ldots, x_{-1}, y_{-1}, x_0, y_0) \in \mathbf{X}^\infty$.
FOR $n = 1, 2, \ldots$:
  Environment announces $x_n \in \mathbf{X}$.
  Predictor announces $\gamma_n \in \Gamma$.
  Environment announces $y_n \in \mathbf{Y}$.
END FOR.

x_n: signal; γ_n: prediction, y_n: observation.

$\lambda : \Gamma \times \mathbf{Y} \to \mathbb{R}$: loss function.

Traditionally in learning theory: no infinite past.
Modeled in our framework by:

- extending the signal and observation space by new elements $? \in \mathbf{X}$ and $? \in \mathbf{Y}$,
- defining $\lambda(\gamma, ?)$ arbitrarily,
- making Environment announce the infinite past $(\ldots, x_{-1}, y_{-1}, x_0, y_0) = (\ldots, ?, ?, ?, ?)$ and refrain from announcing $x_n = ?$ or $y_n = ?$ afterwards.

Stationary strategies

Prediction strategy = Predictor's strategy. Mathematically: a function $D : (\mathbf{X} \times \mathbf{Y})^\infty \times \mathbf{X} \times \{1, 2, \ldots\} \to \Gamma$.

Natural class of strategies: the stationary prediction strategies $D : (\mathbf{X} \times \mathbf{Y})^\infty \times \mathbf{X} \to \Gamma$.

Examples of stationary strategies

Prediction rules: strategies $D : \mathbf{X} \to \Gamma$. No infinite past needed.

Markov strategies of order k: $D : (\mathbf{X} \times \mathbf{Y})^k \times \mathbf{X} \to \Gamma$; only finite past is needed (or no past, if we do not have to predict from the very beginning).

Compact deterministic case: definition

A prediction strategy is CS universal for a loss function λ if its predictions γ_n satisfy

$$\limsup_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^{N}\lambda(\gamma_n,y_n)-\frac{1}{N}\sum_{n=1}^{N}\lambda\Big(D(\ldots,x_{n-1},y_{n-1},x_n),y_n\Big)\right)\le 0$$

for any continuous stationary prediction strategy D and any biinfinite sequence $\ldots,x_{-1},y_{-1},x_0,y_0,x_1,y_1,\ldots$.

Compact deterministic case: result

Proposition 1 Suppose $\mathbf{X}$ and $\mathbf{Y}$ are compact metric spaces, Γ is a compact convex subset of a Banach space, and the loss function $\lambda(\gamma, y)$ is continuous in (γ, y) and convex in the variable $\gamma \in \Gamma$. There exists a CS universal prediction strategy.

Construction: Kalnishkan and Vyugin's Weak Aggregating Algorithm, applied to a dense subset of $\Gamma^{(\mathbf{X}\times\mathbf{Y})^\infty\times\mathbf{X}}$.

Compact randomized case: definitions 1

When the loss function $\lambda(\gamma, y)$ is not convex in γ: the conclusion of Proposition 1 becomes false if the convexity requirement is removed (follows from, e.g., Kalnishkan and Vyugin's results).

A randomized prediction strategy is a function $D : (\mathbf{X} \times \mathbf{Y})^\infty \times \mathbf{X} \times \{1, 2, \ldots\} \to \mathcal{P}(\Gamma)$. A stationary randomized prediction strategy is a function $D : (\mathbf{X} \times \mathbf{Y})^\infty \times \mathbf{X} \to \mathcal{P}(\Gamma)$.

Compact randomized case: definitions 2

$\mathcal{P}(\Gamma)$ is always equipped with the topology of weak convergence.

A randomized prediction strategy outputting γ_n is CS universal for a loss function λ if, for any continuous stationary randomized prediction strategy D and any biinfinite $\ldots, x_{-1}, y_{-1}, x_0, y_0, x_1, y_1, \ldots$,

$$\limsup_{N\to\infty} \left(\frac{1}{N} \sum_{n=1}^{N} \lambda(g_n, y_n) - \frac{1}{N} \sum_{n=1}^{N} \lambda(d_n, y_n) \right) \le 0 \text{ a.s.},$$

where $g_1, g_2, \ldots, d_1, d_2, \ldots$ are independent random variables distributed as

$$\begin{aligned} &g_n \sim \gamma_n, \\ &d_n \sim D(\ldots, x_{n-1}, y_{n-1}, x_n), \qquad n = 1, 2, \ldots . \end{aligned}$$

Compact randomized case: result

Proposition Let $\mathbf{X}$, Γ, and $\mathbf{Y}$ be compact metric spaces and λ be a continuous loss function. There exists a CS universal randomized prediction strategy.

Deterministic case: definitions 1

A set in a topological space is precompact if its closure is compact. In Euclidean spaces, precompactness = boundedness.

A prediction strategy outputting $\gamma_n \in \Gamma$ is CS universal for a loss function λ if, for any continuous stationary prediction strategy D and for any biinfinite $\ldots, x_{-1}, y_{-1}, x_0, y_0, x_1, y_1, \ldots$,

$$\Big(\{\ldots, x_{-1}, x_0, x_1, \ldots\} \text{ and } \{\ldots, y_{-1}, y_0, y_1, \ldots\} \text{ are precompact}\Big)$$
$$\Longrightarrow \limsup_{N\to\infty}\left(\frac{1}{N}\sum_{n=1}^{N}\lambda(\gamma_n, y_n) - \frac{1}{N}\sum_{n=1}^{N}\lambda\Big(D(\ldots, x_{n-1}, y_{n-1}, x_n), y_n\Big)\right) \le 0.$$

Deterministic case: definitions 2

The loss function λ is large at infinity if, for all $y^* \in \mathbf{Y}$,

$$\lim_{\substack{y \to y^* \\ \gamma \to \infty}} \lambda(\gamma, y) = \infty$$

(in the sense that for each constant M there exists a neighborhood $O_{y^*} \ni y^*$ and compact $C \subseteq \Gamma$ such that $\lambda\left(\Gamma \setminus C, O_{y^*}\right) \subseteq (M, \infty)$). Intuitively: faraway $\gamma \in \Gamma$ should be poor predictions for nearby $y \in \mathbf{Y}$.

Deterministic case: result

Theorem 1 Suppose $\mathbf{X}$ and $\mathbf{Y}$ are locally compact metric spaces, Γ is a convex subset of a Banach space, and the loss function $\lambda(\gamma, y)$ is continuous, large at infinity, and convex in the variable $\gamma \in \Gamma$. There exists a CS universal prediction strategy.

Simple example: $\mathbf{X} = \mathbb{R}^K$, $\Gamma = \mathbf{Y} = \mathbb{R}^L$, and $\lambda(\gamma, y) := \|y - \gamma\|$ satisfy the conditions of the theorem.

Randomized case: definitions

We say that a randomized prediction strategy outputting randomized predictions $\gamma_n \in \mathcal{P}(\Gamma)$ is CS universal if, for any continuous stationary randomized prediction strategy D and for any biinfinite $\ldots, x_{-1}, y_{-1}, x_0, y_0, x_1, y_1, \ldots$,

$$\left(\{\ldots, x_{-1}, x_0, x_1, \ldots\} \text{ and } \{\ldots, y_{-1}, y_0, y_1, \ldots\} \text{ are precompact}\right)$$
$$\Longrightarrow \left(\limsup_{N\to\infty} \left(\frac{1}{N}\sum_{n=1}^{N} \lambda(g_n, y_n) - \frac{1}{N}\sum_{n=1}^{N} \lambda(d_n, y_n)\right) \le 0 \text{ a.s.}\right),$$

where $g_1, g_2, \ldots, d_1, d_2, \ldots$ are independent random variables

$$g_n \sim \gamma_n,$$
$$d_n \sim D(\ldots, x_{n-1}, y_{n-1}, x_n), \qquad n = 1, 2, \ldots .$$

Randomized case: result

Theorem 2 Let $\mathbf{X}$ and $\mathbf{Y}$ be locally compact metric spaces, Γ be a metric space, and λ be a continuous and large at infinity loss function. There exists a CS universal randomized prediction strategy.

Theorems 1 and 2 contain Propositions 1 and 2, respectively, as special cases: in the compact case the condition that the loss function should be large at infinity is satisfied automatically.

Idea of the proof of Theorems 1 and 2

When $\mathbf{X}$ and $\mathbf{Y}$ are Euclidean spaces: a version of the "doubling trick".

Arbitrary locally compact spaces: topological version of the "doubling trick".

Game of removal $G(X)$
(X is a topological space):

FOR $n = 1, 2, \ldots$:
    Remover announces compact $K_n \subseteq X$.
    Evader announces $p_n \notin K_n$.
END FOR.

Winner: Evader if the set $\{p_1, p_2, \ldots\}$ is precompact;
    Remover otherwise.

Topological lemma

Intuitively, the goal of Evader is to avoid being removed to the infinity.

Lemma [Gruenhage, 2006] Remover has a winning strategy in $G(X)$ if X is a locally compact and paracompact space.

Idea of the proof of Theorems 1 and 2, cont.

Assuming that Remover outputs rectangles in $G(\mathbf{X} \times \mathbf{Y})$ and plays a winning strategy:

1. $k := 1$ and $n := 1$;
2. Remover makes his next (first if $k = 1$) move in $G(\mathbf{X} \times \mathbf{Y})$, outputting $(\mathbf{X}_k, \mathbf{Y}_k)$;
3. Predictor plays his winning strategy in the game of prediction with the signal space $\mathbf{X}_k$ and observation space $\mathbf{Y}_k$ until $(x_n, y_n) \notin (\mathbf{X}_k, \mathbf{Y}_k)$ (forever if this never happens);
4. $k := k + 1$; go to 2.

Natural next step

Obtain non-asymptotic versions of these results.

Thank you for your attention.