Predictive complexity, randomness and information

Volodya Vovk

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

vovk@cs.rhul.ac.uk

Authors of the results: mainly Vladimir & Michael Vyugin, Yuri Kalnishkan.

This talk: limits of prediction. “Prediction”: coding (Kolmogorov complexity), investing in a stock market, prediction.

What predictive performance is achievable if you do not worry about computational resources?

This talk: 3 closely related topics

- predictive complexity
- predictive randomness
- predictive information

Games of prediction

FOR $i = 1, 2, \ldots$ Forecaster

(1) chooses prediction $p_i \in \mathbf{P}$

(2) observes actual outcome $y_i \in \mathbf{Y}$

(3) suffers loss $\lambda(p_i, y_i)$

END FOR

λ: computable.

Example: square-loss game;

$$\mathbf{Y} = \{0, 1\},\ \mathbf{P} = [0, 1],$$

$$\lambda(p, y) = (p - y)^2.$$

Predictive complexity

If S is a prediction strategy, $\mathrm{Loss}_S(z)$ is the loss incurred by Forecaster who follows S after Nature chooses $z = (y_1, \ldots, y_n)$. We are interested in $\mathrm{Loss}_S(z)$ (loss process of S) for the best computable S. Such S does not exist!

$L : \mathbf{Y}^* \to \mathbb{R}$ is a measure of predictive complexity if:

- L is a superloss process: $L(\Box) = 0$ and

$$\forall z \, \exists p \, \forall y : L(z * y) \geq L(z) + \lambda(p, y);$$

- L is semicomputable from above: there exists a computable sequence of computable functions f_k such that $L = \inf_k f_k$.

A smallest, to within an additive constant, measure $\mathcal{K}$ of predictive complexity is universal;

$$\forall L\,\exists C\,\forall z : \mathcal{K}(z) \le L(z) + C.$$

If it exists: predictive complexity.

Theorem 1. Predictive complexity exists if and only if the prediction game is mixable.

The square-loss game is mixable.

In the binary ($\mathbf{Y} = \{0, 1\}$) case: Canonical representation of a game:

$$\{(\lambda(p, 0), \lambda(p, 1)) : p \in \mathbf{P}\} \subseteq [0, \infty]^2.$$

For an exponential learning rate $\beta = e^{-\eta}$, the exponential representation:

$$\left\{\left(\beta^{\lambda(p,0)}, \beta^{\lambda(p,1)}\right) : p \in \mathbf{P}\right\} \subseteq [0, 1]^2.$$

A game is mixable if for some $\beta \in (0, 1)$ the exponential representation is concave.

- Predictive complexity exists for mixable games: simple (you can mix in the exponential picture).
- If predictive complexity exists, the game is mixable: more complicated.

Log-loss game:

$$\mathbf{Y} = \{0, 1\},\ \mathbf{P} = [0, 1],$$

$$\lambda(p, y) = \begin{cases} -\log p & \text{if } y = 1 \\ -\log(1-p) & \text{if } y = 0 \end{cases}$$

Game of coding (or probability forecasting).

This is what becomes with

$$\exists p\,\forall y: L(z*y) \ge L(z) + \lambda(p,y)$$

in the log-loss case:

$\exists p: L(z{*}0) \ge L(z) - \log(1-p) \,\&\, L(z{*}1) \ge L(z) - \log p$;

$\exists p: 2^{-L(z*0)} \le (1-p)2^{-L(z)} \,\&\, 2^{-L(z*1)} \le p2^{-L(z)}$;

$$2^{-L(z*0)} + 2^{-L(z*1)} \le 2^{-L(z)}.$$

Therefore:

$$\mathcal{K}(z) =^{+} \mathit{KM}(z) := -\log M(z);$$

close to $K(z)$ and $C(z)$.

The game is obviously mixable.

Cover's game with 2 stocks (0 and 1):

$$\mathbf{Y} = [0, \infty)^2,\ \mathbf{P} = [0, 1],$$

$$\lambda(p, (y_0, y_1)) = -\log(p y_1 + (1-p) y_0).$$

The game is mixable and $\mathcal{K}^{\text{Cover}}$ contains $\mathcal{K}^{\log}$, e.g.,

$$\mathcal{K}^{\log}(001110) =$$

$$\mathcal{K}^{\text{Cover}}((1,0),(1,0),(0,1),(0,1),(0,1),(1,0)).$$

Long investment game.

Games for which the predictive complexity does not exist: the simple prediction game

$$\mathbf{Y} = \{0, 1\},\ \mathbf{P} = \{0, 1\},$$

$$\lambda(p, y) = |p - y|;$$

the absolute-loss game

$$\mathbf{Y} = \{0, 1\},\ \mathbf{P} = [0, 1],$$

$$\lambda(p, y) = |p - y|$$

(the expected loss in the simple prediction game).

Now assume $\lambda \geq 0$.

In the absolute-loss game,

$$\exists \mathcal{K} \, \forall L \, \exists C \, \forall z : \mathcal{K}(z) \le L(z) + C\sqrt{n},$$

where $n = l(z)$ is the length of z.

In the simple prediction game,

$$\exists \mathcal{K} \, \forall L \, \exists C \, \forall z : \mathcal{K}(z) \le CL(z).$$

Hypothesis: In the bounded binary case, predictive complexity exists with accuracy $\cdots + O(1)$ (mixable games), $\cdots + O(n^{1/2})$, or $O(\cdots)$; no other accuracy is possible.

Open problem: Find a general algorithm determining the accuracy with which predictive complexity exists.

Predictive randomness

Assume predictive complexity exists.

A finite sequence $x \in Y^n$ is C-random w.r. to a computable prediction strategy S (or S is C-efficient for x) if

$$\mathcal{K}(x \mid n) \geq \mathsf{Loss}_S(x) - C.$$

(Maybe $\mathcal{K}(x)$.) For the log-loss game: coincides with Kolmogorov randomness.

Suppose the CR is symmetric and $(B, B) \in$ CR. Trivial upper bound: $\mathcal{K}(z) \leq Bl(z)$.

Theorem 2. If CR is smooth near (B, B),

$$\sup_{n,m} \frac{\#\{z \in \{0,1\}^n : \mathcal{K}(z \mid m) \le Bn - m\}}{2^n \beta_0^m} \le 1$$

and

$$\inf_{m} \varliminf_{n\to\infty} \frac{\#\{z \in \{0,1\}^n : \mathcal{K}(z \mid m) \le Bn - m\}}{2^n \beta_0^m} > 0,$$

where β_0 is the smallest β such that the β-exponential representation lies below $x + y = 2\beta^B$.

Problem: Study unpredictability for arbitrary loss functions, probability distributions and predictive strategies.

Predictive information:

$$I(a : y_1 \ldots y_n) := \mathcal{K}(y_1 \ldots y_n) - \mathcal{K}(y_1 \ldots y_n \mid a).$$

E.g.: how much message a increases the capital that can be made investing in a stock market represented by the sequence of stock prices $y_1 \ldots y_n$?

Snooping curve:

$$S_z(\alpha) := \max_{a:l(a)\le\alpha} I(a : z).$$

E.g.: the maximal value of α bits of insider information.

Problem. Consider the binary case. Let f be the scaled version of S,

$$f_z(t) := \frac{S_z(nt)}{n},\ t \in [0,1].$$

What are the limiting points of sequences of functions f_{z_i}, $i = 1, 2, \ldots$, $l(z_i) = i$?

Structure function I:

$$H_z(\alpha) := \min_{a : l(a) \le \alpha} \mathcal{K}(z \mid a).$$

It is clear that:

$$H_z(\alpha) = \mathcal{K}(z) - S_z(\alpha),$$

$$H_z(\alpha) = \inf_{L : K(L) \le \alpha} L(z) + O(1),$$

L ranging over upper semicomputable superloss processes.

Structure function II:

$$h_z(\alpha) := \inf_{L : K(L) \le \alpha} L(z),$$

L ranging over computable loss processes.

Equivalently:

$$h_z(\alpha) := \inf_{P : K(P) \le \alpha} \mathrm{Loss}_P(z),$$

P ranging over computable prediction strategies.

Theorem 3. For unbounded monotonic non-decreasing integer-valued $0 < \mu_n < n$ and $\nu_n = o(n)$, define the scaled version

$$f_z(t) := \frac{h_z(\nu_n t)}{\mu_n},\ t \in [0,1].$$

Each non-decreasing $F \in L_\infty[0,1]$ is a limit point (in L_∞) of normalized structure functions.

Problem: Make the vertical and horizontal scale of the same order of magnitude.

Done by Levin and Vereshchagin&Vitanyi for $\mu_n = \nu_n = n$ and the log-loss game.