Universal Prediction of Selected Bits

Tor Lattimore[1]   Marcus Hutter[1,2]   Vaibhav Gavane[3]

October 7, 2011

[1]Australian National University
[2]ETH Zürich
[3]VIT University

The Problem

Can you predict the next bit of each sequence?

111111111111111111111111111111111

101010101010101010101010101010101

111100110011001100110000111100110

1111001100110011001100001111001100

- The first two are easy. The third can be predicted by noticing bits always come in pairs. The fourth is hard.
- We study sequence prediction where it is only reasonable to predict selected bits, as in the 3rd line above.
- Interesting because various learning problems have this structure. For example, online classification.

Formal Setup

Let $\mu : \{0,1\}^* \to [0,1]$ represent a measure on $\{0,1\}^\infty$ where

- $\mu(x)$ is the probability that an infinite string sampled from μ begins with x.
- $\mu(b|x) := \mu(xb)/\mu(x)$ is the probability of seeing $b \in \{0,1\}$ having already seen x.

Let ω be sampled from an unknown μ. The goal is to learn the measure μ sequentially from an increasing prefix of ω.

Solomonoff Induction

Let $\{\nu_1, \nu_2, \cdots\}$ be a set containing all enumerable semi-measures.

$$\mathbf{M}(x) := \sum_{\nu_i} w_{\nu_i} \nu_i(x) \qquad w_{\nu_i} > 0 \qquad \sum_{\nu_i} w_{\nu_i} \leq 1$$

- **M** is universal in that it dominates every enumerable semi-measure.

$$\mathbf{M}(x) > c_\mu \mu(x), \quad \forall \mu$$

- **M** isn't computable, but is enumerable. There exists a recursive $\phi(x, t)$ such that

$$\phi(x, t+1) \geq \phi(x, t) \qquad \lim_{t \to \infty} \phi(x, t) = \mathbf{M}(x)$$

- **M** is not a proper measure.

Solomonoff Induction

Theorem (Solomonoff, 1965)

Let μ be any computable measure and $S(\omega)$ be defined by

$$S(\omega) := \sum_{t=1}^{\infty} \sum_{b \in \{0,1\}} [\mathbf{M}(b|\omega_{<t}) - \mu(b|\omega_{<t})]^2$$

then

$$\mathbf{E}_\mu[S] < K(\mu) \log 2 < \infty$$

Corollary

Let ω be any computable infinite binary string then

$$\sum_{t=1}^{\infty} [\mathbf{M}(\omega_n|\omega_{<n}) - 1]^2 < \infty, \qquad \lim_{t\to\infty} \mathbf{M}(\omega_n|\omega_{<n}) = 1$$

Therefore **M** is universal predictor for computable sequences.

Solomonoff Induction

Definition (Computable Predictor)

A *computable predictor* is a totally recursive function $f : \{0,1\}^* \to \{0,1,\epsilon\}$ where $f(x)$ represents the prediction of f for the next bit. If $f(x) = \epsilon$ then f chooses to make no prediction.

Example

The following function predicts even bits equal to preceding odd bits and doesn't try to predict odd bits.

$$f(x) = \begin{cases} x_n & \text{if } n \text{ odd} \\ \epsilon & \text{otherwise} \end{cases}$$

Remark. It's important that f is totally recursive. It must always halt, even if it doesn't wish to make a prediction.

Prediction of Selected Bits

Definition (Normalisation)

Recall $\mathbf{M}$ is not a proper measure. Solomonoff normalised it to $\mathbf{M}_{norm}$.

$$\mathbf{M}_{norm}(\epsilon) := 1 \qquad \mathbf{M}_{norm}(xb) = \mathbf{M}_{norm}(x) \left[\frac{\mathbf{M}(xb)}{\mathbf{M}(x0) + \mathbf{M}(x1)} \right]$$

- No longer enumerable, only approximable.
- Strictly larger than $\mathbf{M}$.
- Natural, but not unique.

Theorem (Positive Result)

Let f be a computable predictor and ω be an infinite string such that $f(\omega_{<n}) = \omega_n$ for all n where $f(\omega_{<n}) \neq \epsilon$. If $f(\omega_{<n_i}) \neq \epsilon$ for an infinite sequence $n_1, n_2, \cdots$, then

$$\lim_{i \to \infty} \mathbf{M}_{norm}(\omega_{n_i} | \omega_{<n_i}) = 1$$

Prediction of Selected Bits

The theorem implies that $\mathbf{M}_{norm}$ successfully predicts selected bits as well as any computable predictor.

Examples

The following can be predicted by $\mathbf{M}_{norm}$.

- 11110011001100110011000011110011 0
- Computable subsequences of arbitrary noise.
- B0F0Z0N0X0D0E1J0P0R0U1A1S0L0I1T0H0E1U

  0 for vowels, 1 for consanants
- More general deterministic classification.

Result only applies to deterministically generated subsequences. The stochastic case is still an open question.

Failure to Predict Selected Bits

Theorem (Negative Result)

There exists an infinite binary string ω *with* $\omega_{2t} = \omega_{2t-1}$ *for all* $t \geq 1$ *such that*

$$\liminf_{t\to\infty} \mathbf{M}(\omega_{2t}|\omega_{<2t}) < 1.$$

Not all bad though. Possible to show that

$$\mathbf{M}(\omega_{2t}|\omega_{<2t}) > c, \forall t \qquad \text{and} \qquad \lim_{t\to\infty} \mathbf{M}(\neg\omega_{2t}|\omega_{<2t}) = 0.$$

Remarks

- Unusually, there is a "practical" difference between $\mathbf{M}_{norm}$ and $\mathbf{M}$.
- For computable ω, $\mathbf{M}$ makes at most $Km(\omega)\log 2$ mistakes. It is likely no such result is possible in the case of selected bits.
- It's possible to generalise the definition of a computable predictor by only insisting that it is computable on prefixes of ω, not necessarily all $x \in \{0,1\}^*$.
- The disadvantage of $\mathbf{M}_{norm}$ is that it is not enumerable. This is countered by noting that the conditional distributions of both $\mathbf{M}$ and $\mathbf{M}_{norm}$ are only approximable anyway.
- The ω in the negative result is not Martin-Löf random with respect to any computable measure.
- We only consider deterministic sub-patterns. The stochastic case might be much harder.
- None of this can be computed, but may be approximated.

Proof Intuition (positive result)

Theorem (Lempp, Miller, Ng and Turetsky)

For any $\omega \in \{0,1\}^\infty$, $\lim_{t\to\infty} \mathbf{m}(\omega_{<t})/\mathbf{M}(\omega_{<t}) = 0$.

For the positive result show $\mathbf{M}(\omega_{<t}\neg\omega_t) \overset{\times}{=} \mathbf{m}(\omega_{<t}\neg\omega_t)$. Therefore

$$\lim_{t\to\infty} \frac{\mathbf{M}(\omega_{<t}\neg\omega_t)}{\mathbf{M}(\omega_{<t}\omega_t)} = 0$$

Normalising ensures

$$\lim_{t\to\infty} \mathbf{M}_{norm}(\omega_t|\omega_{<t}) + \mathbf{M}_{norm}(\neg\omega_t|\omega_{<t}) = 1,$$

which implies

$$\lim_{t\to\infty} \mathbf{M}_{norm}(\omega_t|\omega_{<t}) = 1.$$

Proof Intuition (negative result)

It is well known that **M** is not a proper measure.

The idea is to show **M** can fail to converge to a proper measure on the even bits of some infinite sequence ω with $\omega_{2t} = \omega_{2t-1}$.

Not as easy as it seems since

1. The set of the sequences on which **M** does not converge to a proper measure has measure zero w.r.t any computable measure.
2. There exists a $c > 0$ such that if ω satisfies $\omega_{2t} = \omega_{2t-1}$ then $\inf \mathbf{M}(\omega_{2t}|\omega_{<2t}) > c$.