Introduction to Universal Artificial Intelligence

Variants and applications of AIXI

Pierre-Alexandre MURENA

January 2020

Outline

Universal lower semicomputable semimeasures

Definition

A discrete semimeasure is a function $P : \mathbb{N} \to \mathbb{R}$ that satisfies $\sum_{x \in \mathbb{N}} P(x) \leq 1$. It is a probability measure if equality holds.

Definition

Let $\mathcal{M}$ be a class of discrete semimeasures. A semimeasure $P_0 \in \mathcal{M}$ is universal for $\mathcal{M}$ if for all $P \in \mathcal{M}$, there exists a constant c_P such that for all $x \in \mathbb{N}$, we have $c_P P_0(x) \geq P(x)$, where c_P does not depend on x.

Theorem

There is a universal lower semicomputable discrete semimeasure $\mathbf{m} = \sum_j \alpha(j) P_j$. It is possible to choose $\alpha(j) = 2^{-K(j) - O(1)}$.

Two other semi-measures

- The universal a priori probability on the positive integers is defined as

$$Q_U(x) = \sum_{U(p)=x} 2^{-l(p)}$$

where U is the prefix reference TM used te define complexity.

- The algorithmic probability $R(x)$ is defined as $R(x) = 2^{-K(x)}$.

Theorem (Coding theorem)

There is a constant c such that for every x:

$$\log \frac{1}{\mathbf{m}(x)} = \log \frac{1}{Q_U(x)} = K(x)$$

with equality up to the additive constant c.

Predictive convergence of M

We consider a binary sequence $x_1, x_2, \ldots$ and the problem of sequence continuation.
We consider the following distribution, called *Solomonoff distribution*:

$$\mathbf{M}(x) = \sum_{p:U(p)=x*} 2^{-l(p)}$$

Theorem

If observations $x_1, x_2, x_3 \ldots$ are drawn from a distribution μ (unknown), then:

$$\sum_{t=1}^{\infty} \sum_{x_{<t} \in \mathbb{B}^{t-1}} \mu(x_{<t}) \left(\mathbf{M}(0|x_{<t}) - \mu(0|x_{<t})\right)^2 < \frac{1}{2} \ln(2) K(\mu) < \infty$$

Sequence prediction: General case

We now consider the case of sequence prediction for any finite alphabet $\mathcal{X}$ (not necessarily binary). We consider a class $\mathcal{M} = \{\nu_1, \nu_2, \dots\}$ of candidate hypotheses, and we suppose $\mu \in \mathcal{M}$.

We define a weighted average

$$\xi_{\mathcal{M}}(x_{1:n}) = \sum_{\nu \in \mathcal{M}} w_\nu \nu(x_{1:n})$$

with $\sum_{\nu \in \mathcal{M}} w_\nu = 1$ and $w_\nu < 0$.

Convergence of ξ to μ

$$(i) \sum_{t=1}^{n} \mathbb{E}\left[\sum_{x'_t} \left(\mu(x'_t|x_{<t}) - \xi(x'_t|x_{<t})\right)^2\right] = S_n \leq D_n \leq \ln w_\mu^{-1} < \infty$$

$$(ii) \sum_{x'_t} \left(\mu(x'_t|x_{<t}) - \xi(x'_t|x_{<t})\right)^2 = s_t(x_{<t}) \leq d_t(x_t) \to 0 \text{ with } \mu \text{ proba } 1$$

$$(iii) \xi(x'_t|x_{<t}) - \mu(x'_t|x_{<t}) \to 0 \text{ with } \mu \text{ proba } 1 \text{ for all } x'_t$$

$$(iv) \sum_{t=1}^{n} \mathbb{E}\left[\left(\sqrt{\frac{\xi(x'_t|x_{<t})}{\mu(x'_t|x_{<t})}} - 1\right)^2\right] \leq H_n \leq D_n < \infty$$

$$(v) \frac{\xi(x'_t|x_{<t})}{\mu(x'_t|x_{<t})} \to 1 \text{ with } \mu \text{ proba } 1$$

Pareto optimality of ξ

Definition

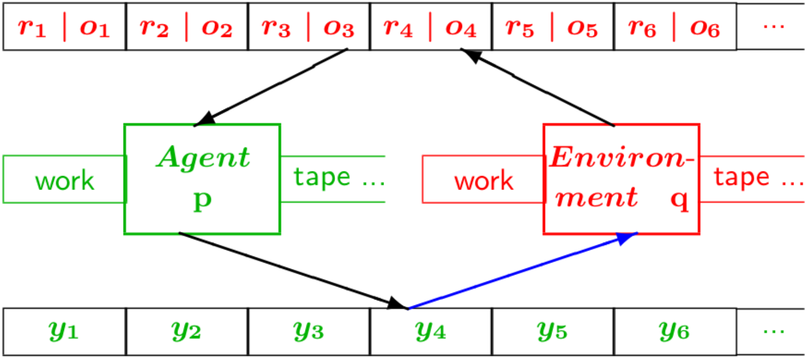
Let $F(\mu, \rho)$ be any performance measure of ρ relative to μ. The universal prior ξ is called pareto optimal wrt F if there is no ρ with $F(\nu, \rho) \leq F(\nu, \xi)$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν.

Theorem

A prior ξ is Pareto optimal wrt the instantaneous and total squared distances s_t and S_n, entropy distances d_t and D_n, errors e_t and E_n and loss l_t and L_n.

Chronological Turing Machine

When the agent's and environment's policies p and q are computable, one can summarize the model as follows:



Value function

- In deterministic environments:

$$V_{km}^{pq} = \sum_{i=k}^{m} r(x_i^{pq})$$

- In probabilistic environments:

$$V_{km}^{p\mu}(\dot{y}\dot{x}_{<k}) = \sum_{q:q(\dot{y}_{<k})=\dot{x}_{<k}} \mu(q) V_{km}^{pq}$$

- The AIμ agent picks action $\dot{y}_k$ according to:

$$\dot{y}_k = \arg\max_{y_k} \max_{p:p(\dot{x}_{<k})=\dot{y}_{<k}y_k} V_{km_k}^{p\mu}(\dot{y}\dot{x}_{<k})$$

Two (equivalent) formulations of the problem

Functional form of AIμ :

$$\dot{y}_k = \arg\max_{y_k} \max_{p : p(\dot{x}_{<k}) = \dot{y}_{<k} y_k} V^{p\mu}_{km_k}(\dot{y}\dot{x}_{<k})$$

Iterative form of AIμ :

$$\dot{y}_k = \arg\max_{y_k} \sum_{x_k} \max_{y_{k+1}} \sum_{x_{k+1}} \ldots \max_{y_m} \sum_{x_m} \left(r(x_k) + \ldots + r(x_m)\right) \mu(x_{k:m} | \dot{y}\dot{x}_{<k} y_{k:m})$$

Equivalent definitions of the AIXI model

Functional form of AIXI: Let $M(q) = 2^{-l(q)}$:

$$\dot{y}_k^{AIXI} = \arg\max_{y_k} \max_{p:p(\dot{x}_{<k})=\dot{y}_{<k}y_k} \sum_{q:q(\dot{y}_{<k})=\dot{x}_{<k}} 2^{-l(q)} V_{km}^{pq}(\dot{y}\dot{x}_{<k})$$

Iterative form of AIXI: Let $M(x_{1:k}|y_{1:k}) = \sum_{q:q(y_{1:k})=x_{1:k}} 2^{-l(q)}$:

$$\dot{y}_k^{AI\xi} = \arg\max_{y_k} \sum_{x_k} \max_{y_{k+1}} \sum_{x_{k+1}} \ldots \max_{y_m} \sum_{x_m} \left(\sum_{i=k}^{m} r(x_i) \right) M(x_{k:m}|\dot{y}\dot{x}_{<k}y_{k:m})$$

Convergence of ξ^{AI} to μ^{AI}: With μ-probability 1:

$$\xi^{AI}(x_{k:m}|x_{<k}y_{1:m}) \to \mu^{AI}(x_{k:m}|x_{<k}y_{1:m})$$

Intelligence order relation

Definition

We call a policy p more or equally inteligent than p' and write $p \succeq p'$ if p yields in any circumstance higher ξ-expected reward than p':

$$\forall k, \forall \dot{y}\dot{x}_{<k}, V_{km}^{p\xi}(\dot{y}\dot{x}_{<k}) \geq V_{km}^{p'\xi}(\dot{y}\dot{x}_{<k})$$

- By definition, $p^\xi \geq p$ for all p: AIXI is the most intelligent agent wrt this order relation.
- Relation $\succeq$ is a universal order relation: it is free of any parameter and any assumption regarding the environment.

Pareto optimality of the Bayes-mixture distribution ξ

We consider here a class $\mathcal{M}$ of possible environments, with $\mu \in \mathcal{M}$, and $w_n u$ such that $\sum_{\nu \in \mathcal{M}} w_\nu = 1$. We define the Bayes-mixture ξ as:

$$\xi(x_{1:m}|y_{1:m}) = \sum_{\nu \in \mathcal{M}} w_\nu \nu(x_{1:m}|y_{1:m})$$

It can be observed that when $\mathcal{M} = \mathcal{M}_u$ and $w_\nu = 2^{-K(\nu)}$, the models AIξ and AIXI are the same!

Definition

A policy $\tilde{p}$ is called Pareto optimal if there is no other policy p with $V_\nu^p \geq V_\nu^{\tilde{p}}$ for all $\nu \in \mathcal{M}$ and strict inequality for at least one ν.

Theorem

Policy p^ξ is Pareto optimal.

Self-optimizing policy

A **self-optimizing policy** $\tilde{p}$ for class $\mathcal{M}$ is an element of $\mathcal{M}$ independent of all μs for which, for horizon $m \to \infty$ and for all μ:

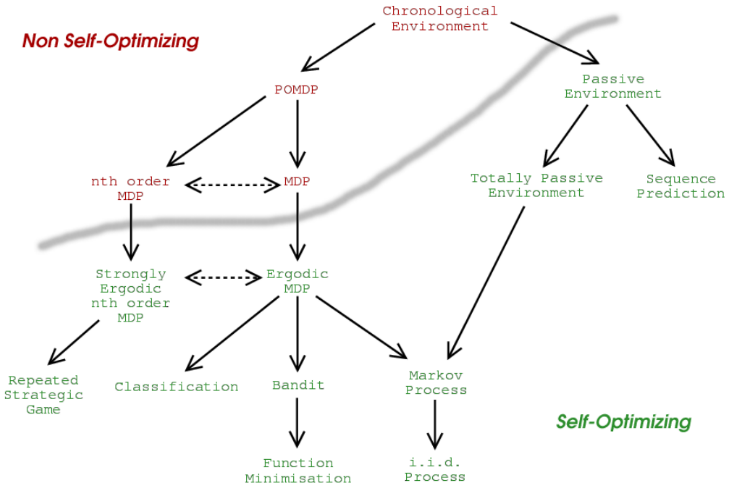
$$\frac{1}{m} V_{1m}^{\tilde{p}\mu} \to \frac{1}{m} V_{1m}^{*\mu}$$

Problem: There is no self-optimizing policy for the class of all chronological environments. (In particular, p^ξ cannot satisfy this property)

Theorem

If $\mathcal{M}$ contains a self-optimizing policy $\tilde{p}$, then p^ξ is also self-optimizing in $\mathcal{M}$.

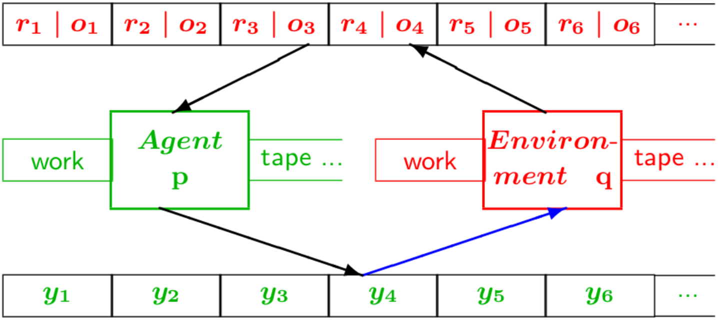
Environments with self-optimizing policies



Outline

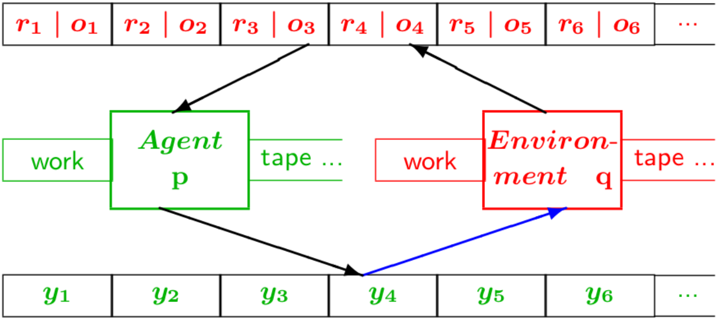
A limitation of AIXI

We used the following framework:



A limitation of AIXI

We used the following framework:



Problem: In general, there is no reward function, and the agent only aims to discover its environment.

Solution: The Knowledge-Seeking Agent (KSA)

The weight $w_k^\nu = w_\nu \frac{\nu(x_{<k}|y_{<k})}{\xi(x_{<k}|y_{<k})}$ is the posterior belief in ν given history $yx_{<k}$. It summarizes the information contained in history $yx_{<k}$

Idea: The changes of w_k quantify the gains of information from one step to the other.

Solution: The Knowledge-Seeking Agent (KSA)

The weight $w_k^\nu = w_\nu \frac{\nu(x_{<k}|y_{<k})}{\xi(x_{<k}|y_{<k})}$ is the posterior belief in ν given history $yx_{<k}$. It summarizes the information contained in history $yx_{<k}$

Idea: The changes of w_k quantify the gains of information from one step to the other.

Formalization: The information gain can be measured by the KL-divergence:

$$r_k = KL(w_{k+1}||w_k)$$

Value function for KSA

Given a discount function γ, we define the value function of KSA as:

$$V_\gamma^\pi(yx_{<k}) = \sum_{t=k}^{\infty} \gamma_t \sum_{\nu \in \mathcal{M}} w_\nu(yx_{<k}) KL(P_\nu^\pi || P_\xi^\pi)$$

and the optimal policy π^* is the policy maximising the falue function:

$$\pi^* = \arg\max_\pi V_\gamma^\pi$$

But does such a policy exist... ?

Value function for KSA

Given a discount function γ, we define the value function of KSA as:

$$V_\gamma^\pi(yx_{<k}) = \sum_{t=k}^{\infty} \gamma_t \sum_{\nu \in \mathcal{M}} w_\nu(yx_{<k}) KL(P_\nu^\pi || P_\xi^\pi)$$

and the optimal policy π^* is the policy maximising the falue function:

$$\pi^* = \arg\max_\pi V_\gamma^\pi$$

But does such a policy exist... ?

Theorem

If $\sum_t \gamma_t < \infty$, then $\sup_\pi V_\gamma^\pi < \infty$ and π^* exists.

On-policy learning, off-policy prediction

Theorem

Let $\mu \in \mathcal{M}$ and γ be a discount vector. If π^* exists, then:

$$\lim_{n\to\infty} \left(\sum_{t=n}^{\infty}\right)^{-1} \mathbb{E}_{\mu}^{\pi^*}\left[\sup_{\pi\in\Pi(yx_{1:n})} \sum_{t=n}^{\infty} \gamma_t KL(P_{\mu}^{\pi}||P_{\xi}^{\pi})\right] = 0$$

Explanation: If the history $yx_{1:\infty}$is generated by π^*, then $P_{\xi}^{\pi}(.|yx_{<n})$ converges in expectation to $P_{\mu}^{\pi}(.|yx_{<n})$.

In other words, the agent learns to predict the counterfactuals "*What would happen if I follow another policy μ instead?*"

KSA for all computable deterministic environments

Suppose that $\mathcal{M}$ is the class of all (semi-)computable environments. As always, we propose to use $w_\nu = 2^{-K(\nu)}$

Proposition

With this choice, we have $V^*_\infty = \infty$.

Remark: A solution to avoid this problem is to choose $w_\nu = 2^{-(1+\epsilon)K(\nu)}$, but:

1. ϵ is an arbitrary parameter!
2. haveing $V^*_\infty < \infty$ is not sufficient to ensure the existence of π^* when $\gamma = \infty$

Outline

Approximations of AIXI

The problem is more difficult than what it seems to be...

see Hutter's slides!

Outline

Where we are

1. ~~Turing machines and computability~~
2. ~~Plain Kolmogorov complexity~~
3. ~~A very little bit of source coding~~
4. ~~Prefix complexity~~
5. ~~Algorithmic probability~~
6. ~~Universal induction~~
7. ~~Universal rational agents~~
8. ~~Approximation of universal rational agents~~

You are free now!

A general model of induction

What is missing?

Deduction			
Type of inference	Specialization	**Framework**	Logical axioms
Assumptions	Non-logical axioms	**Inference rule**	Modus ponens
Results	Theorems	**Universal scheme**	Zermelo-Fraenkel set theory
Universal inference	Universal theorem prover	**Limitation**	Incomplete
In practice	Semi-formal proofs	**Operation**	Proof

Induction			
Type of inference	Generalization	**Framework**	Probability axioms
Assumptions	Prior	**Inference rule**	Bayes rule
Results	Posterior	**Universal scheme**	**Solomonoff probability**
Universal inference	**Universal induction**	**Limitation**	**Incomputability**
In practice	Approximations (?)	**Operation**	Computation

If you are interested...

- Keynote on AIXI approximation:
  `https://www.youtube.com/watch?v=7PFDwh7q3Qs`
- Hutter's interview on intelligence:
  `https://www.youtube.com/watch?v=F2bQ5TSB-cE`
- Hutter's keynote on AGI:
  `https://www.youtube.com/watch?v=YMdFUay0k20`
- Book: *Meta Math!: The Quest for Omega* by Gregory Chaitin

If you are VERY interested

PhD position in Theoretical Computer Science at Higher School of Economics in Moscow, faculty of computer science https://cs.hse.ru/en/about .
Preference is given to students interested in (algorithmic) information theory and related areas such as algorithmic randomness or theoretical foundations of machine learning. Strong students interested in other areas of theoretical computer science are also welcome, for example: algorithms, computational complexity, quantum computation, and more. The student will be supervised by Bruno Bauwens https://www.hse.ru/en/org/persons/160550073 and will be part of the theoretical computer science lab https://cs.hse.ru/en/big-data/tcs-lab/seminars . He is expected to study several areas of theoretical computer science more profoundly and to conduct mathematical research.