Journal of Artificial General Intelligence 1 (2009) 54-61 Submitted 2009-03-14; Revised 2009-09-25

Bias and No Free Lunch in Formal Measures of Intelligence

Bill Hibbard TEST@SSEC.WISC.EDU
University of Wisconsin – Madison
1225 West Dayton Street
Madison, WI 53706

Editor: Catherine Recanati

Abstract

This paper shows that a constraint on universal Turing machines is necessary for Legg's and Hutter's formal measure of intelligence to be unbiased. Their measure, defined in terms of Turing machines, is adapted to finite state machines. A No Free Lunch result is proved for the finite version of the measure.

Keywords: Kolmogorov complexity, no free lunch theorem

1. Introduction

Legg and Hutter have developed a formal mathematical model for defining and measuring the intelligence of agents interacting with environments (Legg and Hutter 2006). Their model includes weighting distributions over time and environments. This paper argues that a constraint on the weighting over environments is required to eliminate bias in the intelligence measure.

The next section of this paper describes Legg's and Hutter's measure and demonstrates the significance of the weighting over environments. Their measure is defined in terms of Turing machines and the third section investigates how the measure can be adapted to a finite model of computing. The fourth section proves an analog of the No Free Lunch Theorem (NFLT) for this finite model.

2. A Formal Measure of Intelligence

Legg and Hutter used reinforcement learning as a framework for defining and measuring intelligence (Legg and Hutter, 2006). In their framework an agent interacts with its environment at a sequence of discrete times, sending action a_i to the environment and receiving observation o_i and reward r_i from the environment at time i. These are members of finite sets A, O and R respectively, where R is a set of rational numbers between 0 and 1. The environment is defined by a probability measure:

$$\mu(o_k r_k \mid o_1 r_1 a_1 \ \ldots \ o_{k-1} r_{k-1} a_{k-1} a_k)$$

and the agent is defined by a probability measure:

$$\pi(a_k \mid o_1 r_1 a_1 \ \ldots \ o_{k-1} r_{k-1} a_{k-1}).$$

The value of agent π in environment μ is defined by the expected value of rewards:

$$V_\mu^\pi = \mathbf{E}(\sum_{i=1}^{\infty} w_i r_i)$$

where the $w_i \geq 0$ are a sequence of weights for future rewards subject to $\sum_{i=1}^{\infty} w_i = 1$ (Legg and Hutter combined the w_i into the r_i). In reinforcement learning the w_i are often taken to be $(1-\gamma)\gamma^{i-1}$ for some $0 < \gamma < 1$. Note $0 \leq V_\mu^\pi \leq 1$.

The intelligence of agent π is defined by a weighted sum of its values over a set E of computable environments. Environments are computed by programs, finite binary strings, on some prefix universal Turing machine (PUTM) U. The weight for $\mu \in E$ is defined in terms of its Kolmogorov complexity:

$$K(\mu) = \min \{ |p| : U(p) \text{ computes } \mu \}$$

where $|p|$ denotes the length of program p. The intelligence of agent π is:

$$V^\pi = \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi.$$

The value of this expression for V^π is between 0 and 1 because of Kraft's Inequality for PUTMs (Li and Vitányi, 1997):

$$\sum_{\mu \in E} 2^{-K(\mu)} \leq 1.$$

Legg and Hutter state that because $K(\mu)$ is independent of the choice of PUTM up to an additive constant that is independent of μ, we can simply pick a PUTM. They do caution that the choice of PUTM can affect the relative intelligence of agents and discuss the possibility of limiting the state-symbol complexity of PUTMs. But as the following proposition illustrates, in order to avoid bias toward specific environments, a constraint on PUTMs is a necessary addition to the definition of their intelligence measure.

Proposition 1. Given $\mu \in E$ and $\varepsilon > 0$ there exists a PUTM U_μ such that for all agents π:

$$V_\mu^\pi / 2 \leq V^\pi < V_\mu^\pi / 2 + \varepsilon$$

where V^π is computed using U_μ.

Proof. Fix a PUTM U_0 that computes environments. Given $\mu \in E$ and $\varepsilon > 0$, fix an integer n such that $2^{-n} < \varepsilon$. Then construct a PUTM U_μ that computes μ given the program "1", fails to halt (alternatively, computes μ) given a program starting with between 1 and n 0's followed by a 1, and computes $U_0(p)$ given a program of n+1 0's followed by p. Now define K using U_μ. Clearly:

$$2^{-K(\mu)} = 1/2$$

And, applying Kraft's Inequality to U_0:

$$\sum_{\mu' \neq \mu} 2^{-K(\mu')} \leq 2^{-n} < \varepsilon.$$

So:

$$V^\pi = V_\mu^\pi / 2 + X$$

Where

$$X = \sum_{\mu' \neq \mu} 2^{-K(\mu')} V_{\mu'}^\pi \text{ and } 0 \leq X < \varepsilon. \quad \square$$

Even if we limit the state-symbol complexity of PUTMs, a small number of environments with short programs may dominate the measured intelligence of agents. The prefix-free encoding of PUTM programs should be designed to prevent any small set of programs from dominating the total weight of the intelligence measure. The following proposition shows how to do this from an arbitrary UTM.

Proposition 2. Given a UTM U that takes all binary strings as programs (i.e., U is not a PUTM) and a positive integer L, U restricted to programs of length at least L is a UTM and these programs can be encoded for a PUTM U'.

Proof. For every U program p such that $|p| < L$, Rice's Theorem (Rice, 1953) says that the property $P(x)$ of U programs, "program x always computes the same values as p", is undecidable. If there were only finitely many such x, then $P(x)$ could be decided by an algorithm that simply compared x to each program in the finite list, so there must be infinitely many. Hence there must be some x computing the same values as p and such that $|x| \geq L$. Thus U restricted to programs with length at least L is still universal.

For each integer $n \geq L$, encode all 2^n U programs of length n by adding a prefix code string C_n consisting of $n–L$ 0's followed by a 1. This gives 2^n program strings of length 2^{2n-L+1} with total weight 2^{L-n-1}. The set of all encoded programs, for $n \geq L$, form a prefix-free code and their total weight is 1. Define a PUTM U' that accepts this set of encoded programs. It strips off the prefix strings C_n and uses U to execute the resulting programs. □

The bias from a small set of environments can be reduced as much as desired in this way, by picking a large L.

The choice of PUTM can affect the relative intelligence of agents. In fact a PUTM can be chosen to produce the counter-intuitive result of an agent with greater measured intelligence than AIXI, which Hutter defined as a maximally intelligent agent (Hutter 2004). The definition of AIXI is quite complex, but the details are not necessary here. The important point is that AIXI uses a PUTM U_{AIXI} to simulate all computable, deterministic environments. At time step k AIXI enumerates all environments that could have produced the interaction history up to time k, then produces the action a_k that maximizes the expected future sum of rewards in interactions with these enumerated environments, where each environment is weighted by its Kolmogorov complexity computed using U_{AIXI}. Because AIXI always chooses the action with maximum expected future sum of rewards, its actions are deterministic unless multiple actions produce the same maximum expected future sum of rewards. The definition of AIXI does not specify how to choose among multiple optimal actions, but that won't matter in the proof of Proposition 3.

As Legg and Hutter state, AIXI has maximal intelligence by their measure, assuming the same PUTM is used to define AIXI and to define the intelligence measure. However, the following proposition shows this is not necessarily the case if these definitions use different PUTMs.

Proposition 3. Assume that $|A| \geq 2$ (i.e., there are at least two possible agent actions) and AIXI is defined using some PUTM U_{AIXI}. Then there exist an environment μ and an agent π such that $V^{\text{AIXI}} < V^{\pi}$, where the intelligence measure V is defined using a PUTM U_{μ} derived from μ according to Proposition 1.

Proof. The action of AIXI at the first time step is a probability distribution P over A (it may be a deterministic distribution, where one action has probability 1 and all others have probability 0). There must be an action $a_1 \in A$ such that the probability $P(a_1) \leq 1/|A|$. Define the action of π at the first time step to be a_1, and define μ to give reward $r_1 = 1$ to action a_1 and reward $r_1 = 0$ to all other actions in the first time step. Note in the definition of environment μ that rewards at every time step are dependent on agent actions at all previous time steps, including the first time step. So define μ to give reward $r_i = 1$ at every time step i, if the agent action was a_1 at the first time step, and to give reward $r_i = 0$ at every time step i, if the agent action was any action other than a_1 at the first time step.

Then the reward to π is 1 at each time step, so $V_{\mu}^{\pi} = 1$ (this is the expected value of the weighted sequence of rewards to π). The expected reward to AIXI is $P(a_1)$ at each time step, so

$V_\mu^{AIXI} = P(a_1)$ (this is the expected value of the weighted sequence of rewards to AIXI). Now apply Proposition 1 to μ and ε to get a PUTM U_μ for an intelligence measure under which:

$$V^{AIXI} < V_\mu^{AIXI} / 2 + \varepsilon = P(a_1) / 2 + \varepsilon \leq 1/(2|A|) + \varepsilon \leq 1/4 + \varepsilon$$

$$1/2 = V_\mu^{\pi} / 2 < V^{\pi}.$$

So, for $\varepsilon < 1/4$, $V^{AIXI} < V^{\pi}$. □

To reiterate, this is only possible because $U_{AIXI} \neq U_\mu$. In the proof, μ and π are designed to conspire to give π higher measured intelligence than AIXI, and hence illustrate the possible pathology of allowing arbitrary PUTMs.

Proposition 3 shows by example that there exist weightings of environments for which some agents have higher measured intelligence than other agents. The construction used in the proof of Proposition 3 can be modified to show the existence of environment weightings for which the difference in measured intelligence between two agents can be arbitrarily close to 1. However, the environment weightings in Proposition 3 are pathological. Proposition 4 will show that for certain environment weightings in a modified, finite version of the intelligence measure, all agents have the same measured intelligence. It is an interesting open question for future research to analyze the distribution of measured intelligence of agents for reasonable environment weightings.

3. A Finite Model

The no free lunch theorem (NFLT) tells us that all optimization algorithms have equal performance when averaged over all finite problems (Wolpert and Macready, 1997). Although the mathematical definition of agents interacting with environments is different from the definition of optimization algorithms interacting with problems, the fact that agents and optimization algorithms are both "trying to do as well as possible" against a set of challenges suggests that there may be a way to reinterpret the NFLT in the context of intelligence measures. That is the goal of this and the next section.

The NFLT is proved for finite problems and encounters difficulties in infinite cases (Auger and Teytaud, 1997), so we will adapt Legg's and Hutter's measure to a finite model. This will lose access to the rich theory of Turing machines, such as the existence of universal Turing machines and the impossibility of solving the halting problem. However, Wang makes a convincing argument that finite and limited resources are an essential component of a definition of intelligence (Wang, 1995). Lloyd estimates that the universe contains no more than 10^{90} bits of information and can have performed no more than 10^{120} elementary operations during its history (Lloyd, 2002), in which case our universe is a finite state machine (FSM) with no more than $2^{(10^{90})}$ states. An intelligence measure based on a finite model is consistent with finite physics and conforms to Wang's argument.

As before, assume the sets A, O and R of actions, observations and rewards are finite and fixed. A FSM is defined by a mapping:

$$f: S(n) \times A \rightarrow S(n) \times O \times R$$

where $S(n)=\{1,2,3,\ldots,n\}$ is a set of states and "1" is the start state (we assume deterministic FSMs so f is single-valued). Letting s_i denote the state at time step i, the timing is such that $f(s_i,a_i) = (s_{i+1},o_i,r_i)$. Because the agent π may be nondeterministic its value in this environment is defined by the expected value of rewards:

$$V_f^{\pi} = \mathbf{E}(\sum_{i=1}^{M(n)} w_{n,i} r_i)$$

where the $w_{n,i} \geq 0.0$ are a sequence of weights for future rewards subject to $\sum_{i=1}^{M(n)} w_{n,i} = 1$ and $M(n)$ is a finite time limit depending on state set size. Note that different state set sizes have different time weights, possibly giving agents more time to learn more complex environments.

Define $F(n)$ as the set of all FSMs with the state set $S(n)$. Define:

$$F = \bigcup_{n=L}^{H} F(n)$$

as the set of all FSMs with state set size between L and H. Define weights W_n such that $\sum_{n=L}^{H} W_n = 1$, and for $f \in F(n)$ define $W(f) = W_n / |F(n)|$. Then $\sum_{f \in F} W(f) = 1$ and we define the intelligence of agent π as:

$$V^{\pi} = \sum_{f \in F} W(f)\, V_f^{\pi}.$$

The lower limit L on state set size is intended to avoid domination of V^{π} by the value of π in a small number of environments. The upper limit H on state size means that intelligence is determined by an agent's value in a finite number of environments. This avoids the necessity for weights to tend toward zero as environment complexity increases. In fact, the weights W_n may be chosen so that more complex environments actually have greater weight than simpler environments.

State is not directly observable so this model counts multiple FSMs with identical behavior. This can be regarded as implicitly weighting behaviors by counting numbers of representations.

4. A No Free Lunch Result

The finite model in the previous section lacks an important hypothesis of the NFLT: that the optimization algorithm never makes the same action more than once. This is necessary to conclude that the ensembles of rewards are independent at different times. The following constraint on the finite model achieves the same result:

Definition. An environment FSM satisfies the No Repeating State Condition (NRSC) if it can never repeat the same state. Such environments must include one or more final states (successor undefined) and a criterion of the NRSC is that every path from the start state to a final state has length $\geq M(n)$, the time limit in the sum for V_f^{π} (this is only possible if $M(n) \leq n$).

Although the NRSC may seem somewhat artificial, it applies in the physical universe because of the second law of thermodynamics (under the reasonable assumption an irreversible process is always occurring somewhere). FSMs satisfying the NRSC are not trivial environments, because our physical universe is not trivial. Also, consider that the state set of our universe is composed of the cross product of the state sets of many subsystems, such as the states of many different stars on an astronomical scale, and the states of individual humans on a social level. For the system as a whole to repeat its state, each subsystem must individually and simultaneously repeat its state. This is so unlikely as to be effectively impossible. Similarly, for complex artificial environments whose state sets are composed of cross products of the state sets of large numbers of subsystems, the NRSC is a natural condition.

Proposition 4. In the finite model defined in the previous section, assume that $M(n) \leq n$ and restrict F to those FSMs satisfying the NRSC. Then for any agent π, $V^{\pi} = (\sum_{r \in R} r) / |R|$, the average reward. Thus all agents have the same measured intelligence.

Proof. Given an agent π, calculate:

$$V^{\pi} = \sum_{f \in F} W(f)\, V_f^{\pi} =$$

$$\sum_{n=L}^{H} \sum_{f \in F(n)} W(f)\, V_f^{\pi} =$$

$$\sum_{n=L}^{H} (W_n / |F(n)|) \sum_{f \in F(n)} V_f^{\pi} =$$

$$\sum_{n=L}^{H} (W_n / |F(n)|) \sum_{f \in F(n)} \mathbf{E}(\sum_{i=1}^{M(n)} w_{n,i}\, r_{f,i}) =$$

$$\sum_{n=L}^{H} (W_n / |F(n)|) \sum_{i=1}^{M(n)} w_{n,i} \sum_{f \in F(n)} \mathbf{E}(r_{f,i}).$$

where $r_{f,i}$ denotes the reward to the agent from environment f at time step i.

To analyze $\sum_{f \in F(n)} \mathbf{E}(r_{f,i})$, define $P(s,a|i,f)$ as the probability that in a time sequence of interactions between agent π and environment f, π makes action a and f is in state s at time step i. Also define $P(r|i,f)$ as the probability that f makes reward r at time step i. Note:

$$\sum_{a \in A} \sum_{s \in S} P(s,a|i,f) = 1 \tag{1}$$

Let f^R denote the R-component of a map $f: S(n) \times A \to S(n) \times O \times R$. For any $s \in S$ and $a \in A$, partition $F(n)$ into the disjoint union $F(n) = \bigcup_{r \in R} F(s,a,r)$ where $F(s,a,r) = \{ f \in F(n) \mid f^R(s,a) = r \}$. Define a deterministic probability:

$$P(r|f,s,a) = 1 \text{ if } f \in F(s,a,r),$$
$$= 0 \text{ otherwise.}$$

Given any two reward values $r_1, r_2 \in R$ (here these do not denote the rewards at the first and second time steps) there is a one-to-one correspondence between $F(s,a,r_1)$ and $F(s,a,r_2)$ as follows: $f_1 \in F(s,a,r_1)$ corresponds with $f_2 \in F(s,a,r_2)$ if $f_1 = f_2$ everywhere except:

$$f_1^R(s,a) = r_1 \neq r_2 = f_2^R(s,a).$$

(Changing a reward value does not affect whether a FSM satisfies the NRSC.) Given such f_1 and f_2 in correspondence, because of the NRSC f_1 and f_2 can only be in state s once, and because they are in correspondence they will interact identically with the agent π before reaching state s. Thus:

$$P(s,a|i,f_1) = P(s,a|i,f_2) \tag{2}$$

Because of the one-to-one correspondence between $F(s,a,r_1)$ and $F(s,a,r_2)$ for any $r_1, r_2 \in R$, and because of equation (2), the value of $\sum_{f \in F(s,a,r)} P(s,a|i,f)$ is independent of r and we denote it by $Q(i,s,a)$. We use this and equation (1) as follows:

$$|F(n)| = \sum_{f \in F(n)} 1 =$$

$$\sum_{f \in F(n)} \sum_{a \in A} \sum_{s \in S} P(s,a|i,f) =$$

$$\sum_{a \in A} \sum_{s \in S} \sum_{f \in F(n)} P(s,a|i,f) =$$

$$\sum_{a \in A} \sum_{s \in S} \sum_{r \in R} \sum_{f \in F(s,a,r)} P(s,a|i,f) =$$

$$\sum_{a \in A} \sum_{s \in S} \sum_{r \in R} Q(i,s,a) =$$

$$\sum_{a \in A} \sum_{s \in S} |R|\, Q(i,s,a).$$

So for any $r \in R$:

$$\sum_{a \in A} \sum_{s \in S} \sum_{f \in F(s,a,r)} P(s,a|i,f) = \tag{3}$$

$$\sum_{a\in A} \sum_{s\in S} Q(i,s,a) =$$

$$|F(n)| / |R|.$$

Now we are ready to evaluate $\sum_{f\in F(n)} \mathbf{E}(r_{f,i})$:

$$\sum_{f\in F(n)} \mathbf{E}(r_{f,i}) =$$

$$\sum_{f\in F(n)} \sum_{r\in R} r\, P(r|i,f) =$$

$$\sum_{f\in F(n)} \sum_{r\in R} r \sum_{a\in A} \sum_{s\in S} P(r|f,s,a)\, P(s,a|i,f) =$$

$$\sum_{r\in R} r \sum_{a\in A} \sum_{s\in S} \sum_{f\in F(n)} P(r|f,s,a)\, P(s,a|i,f) =$$

$$\sum_{r\in R} r \sum_{a\in A} \sum_{s\in S} \sum_{f\in F(s,a,r)} P(s,a|i,f) = \text{(by 3)}$$

$$\sum_{r\in R} r\, |F(n)| / |R| = |F(n)| \,(\sum_{r\in R} r) / |R|.$$

Plugging this back into the expression for V^{π}:

$$V^{\pi} = \sum_{n=L}^{H} (W_n / |F(n)|) \sum_{i=1}^{M(n)} w_{n,i} \sum_{f\in F(n)} \mathbf{E}(r_{f,i}) =$$

$$\sum_{n=L}^{H} (W_n / |F(n)|) \sum_{i=1}^{M(n)} w_{n,i}\, |F(n)| \,(\sum_{r\in R} r) / |R| =$$

$$\sum_{n=L}^{H} (W_n / |F(n)|)\, |F(n)| \,(\sum_{r\in R} r) / |R| =$$

$$(\sum_{r\in R} r) / |R|. \quad \square$$

This proposition satisfies our curiosity about whether and how the NFLT can be reinterpreted in the context of intelligence measures. It also provides evidence of the need for a non-uniform weighting of environments. With infinitely many environments, such as in Legg's and Hutter's intelligence measure, non-uniform weights are inevitable in order to have a finite total weight. But even in a model with finitely many environments, Proposition 4 shows the necessity for non-uniform weights.

By letting $L = H$ in the finite model, Proposition 4 applies to a distribution of environments defined by FSMs with the same state set size.

It would be interesting to construct a PUTM in Legg's and Hutter's model for which all agents have the same measured intelligence within an arbitrarily small ε. It is not difficult to construct a PUTM, somewhat similar to the one defined in the proof of Proposition 1, that gives equal weight to a set of programs defining all FSMs with state set size n satisfying the NRSC, and gives arbitrarily small weight to all other programs. The difficulty is that multiple FSMs will define the same behavior and only one of those FSMs will be counted toward agent intelligence, since Legg's and Hutter's measure sums over environment behaviors rather than over programs. But if their measure had summed over programs, then a PUTM could be constructed for which an analog of Proposition 4 could be proved.

5. Conclusion

Some choices of PUTM can produce extreme bias in Legg's and Hutter's formal measure of intelligence. This bias can be reduced as much as desired by imposing a minimum length limit on programs used to define environments.

According to current physics our universe is a FSM satisfying the NRSC. So it is not unreasonable to measure intelligence using environments defined by FSMs satisfying the NRSC. However, if we measure agent intelligence using a distribution of FSMs satisfying the NRSC in which all FSMs with the same number of states have the same weight, then Proposition 4 shows that all agents have the same measured intelligence. This provides rigorous support that an intelligence measure must based on unequal weighting of environments, such as the weighting based on Kolmogorov complexity used by Legg and Hutter. With an equal weighting of environments, past behavior of environments provides no information about their future behavior. There is a large literature relating to the NFLT, including many papers applying it to various problem areas (http://www.no-free-lunch.org/). This paper applies it to formal measures of intelligence.

References

Auger, A. and O. Teytaud. Continuous lunches are free! *Proceedings of the 9th annual conference on Genetic and evolutionary computation* (*ACM SIGEVO 1997*), pages 916 – 922.

Hutter, M. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin. 2004. 300 pages.

Legg, S. and M. Hutter. Proc. A Formal Measure of Machine Intelligence. *15th Annual Machine Learning Conference of Belgium and The Netherlands* (*Benelearn 2006*), pages 73-80.

Li, M. and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications, 2nd ed.*. Springer, New York, 1997. 637 pages.

Lloyd, S. Computational Capacity of the Universe. Phys.Rev.Lett. 88 (2002) 237901.

Rice, H. G. Classes of Recursively Enumerable Sets and Their Decision Problems. Trans. Amer. Math. Soc. **74**, pages 358-366. 1953.

Wang, P. Non-Axiomatic Reasoning System --- Exploring the essence of intelligence. PhD Dissertation, Indiana University Comp. Sci. Dept. and the Cog. Sci. Program, 1995.

Wolpert, D. and W. Macready, No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation* **1**, 67. 1997.