Towards a Universal Theory of Artificial Intelligence based on Algorithmic Probability and Sequential Decisions

Marcus Hutter

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
IDSIA, Galleria 2, CH-6928 Manno-Lugano, Switzerland
marcus@idsia.ch, http://www.idsia.ch/~marcus

2000 – 2002

Decision Theory	=	Probability + Utility Theory
+		+
Universal Induction	=	Ockham + Epicur + Bayes
=		=
Universal Artificial Intelligence without Parameters		

Table of Contents

Overview

- Decision Theory solves the problem of rational agents in uncertain worlds if the environmental probability distribution is known.
- Solomonoff's theory of Universal Induction solves the problem of sequence prediction for unknown prior distribution.
- We combine both ideas and get a parameterless model of Universal Artificial Intelligence.

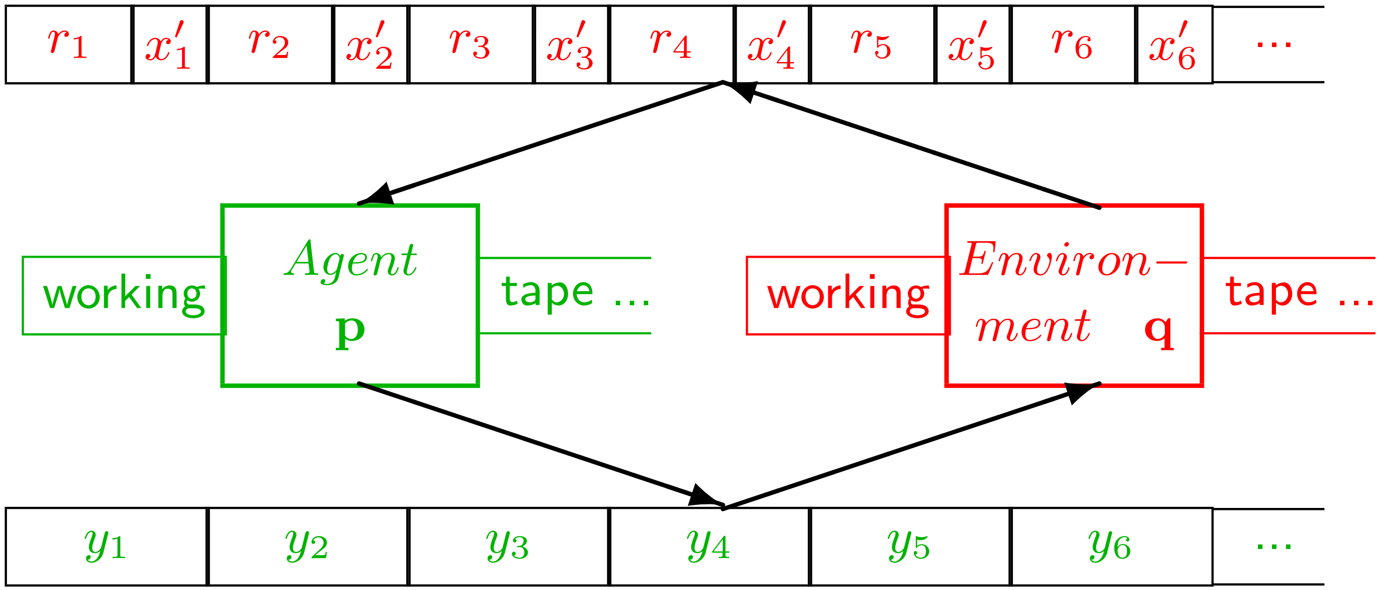
Decision Theory	=	Probability + Utility Theory
+		+
Universal Induction	=	Ockham + Epicur + Bayes
=		=

Universal Artificial Intelligence without Parameters

Preliminary Remarks

- The goal is to mathematically define a unique model superior to any other model in any environment.
- The AIξ model is unique in the sense that it has no parameters which could be adjusted to the actual environment in which it is used.
- In this first step toward a universal theory we are not interested in computational aspects.
- Nevertheless, we are interested in maximizing a utility function, which means to learn in as minimal number of cycles as possible. The interaction cycle is the basic unit, not the computation time per unit.

The Cybernetic or Agent Model



- $p:X^*\to Y^*$ is cybernetic system / agent with chronological function / Turing machine.
- $q:Y^*\to X^*$ is deterministic computable (only partially accessible) chronological environment.

The Agent-Environment Interaction Cycle

for k:=1 to T do

- p thinks/computes/modifies internal state.
- p writes output $y_k \in Y$.
- q reads output y_k.
- q computes/modifies internal state.
- q writes reward/utlitity input $r_k := r(x_k) \in R$.
- q write regular input $x'_k \in X'$.
- p reads input $x_k := r_k x'_k \in X$.

endfor

- T is lifetime of system (total number of cycles).
- Often $R = \{0,1\} = \{bad, good\} = \{error, correct\}$.

Utility Theory for Deterministic Environment

The ($system$,$environment$) pair (p,q) produces the unique I/O sequence

$$\omega(p,q) \;:=\; y_1^{pq}x_1^{pq}y_2^{pq}x_2^{pq}y_3^{pq}x_3^{pq}...$$

Total reward (value) in cycles k to m is defined as

$$V_{km}(p,q) \;:=\; r(x_k^{pq})+...+r(x_m^{pq})$$

Optimal system is system which maximizes total reward

$$p^{best} \;:=\; \underset{p}{\text{maxarg}}\, V_{1T}(p,q)$$

$$\Downarrow$$

$$V_{kT}(p^{best},q)\geq V_{kT}(p,q) \quad \forall p$$

Probabilistic Environment

Replace q by a probability distribution $\mu(q)$ over environments.

The total expected reward in cycles k to m is

$$V_{km}^{\mu}(p|\dot{y}\dot{x}_{<k}) \;:=\; \frac{1}{\mathcal{N}} \sum_{q:q(\dot{y}_{<k})=\dot{x}_{<k}} \mu(q)\cdot V_{km}(p,q)$$

The history is no longer uniquely determined.

$\dot{y}\dot{x}_{<k} := \dot{y}_1\dot{x}_1...\dot{y}_{k-1}\dot{x}_{k-1}$:=actual history.

AIμ maximizes expected future reward by looking $h_k \equiv m_k - k + 1$ cycles ahead (horizon). For $m_k = T$, AIμ is optimal.

$$\dot{y}_k \;:=\; \underset{y_k}{\mathrm{maxarg}} \max_{p:p(\dot{x}_{<k})=\dot{y}_{<k}y_k} V_{km_k}^{\mu}(p|\dot{y}\dot{x}_{<k})$$

Environment responds with $\dot{x}_k$ with probability determined by μ.

This functional form of AIμ is suitable for theoretical considerations. The iterative form (next section) is more suitable for 'practical' purpose.

Iterative AIμ Model

The probability that the environment produces input x_k in cycle k under the condition that the history is $y_1 x_1 ... y_{k-1} x_{k-1} y_k$ is abbreviated by

$$\mu(y\!x_{<k} y\!\underline{x}_k) \;\equiv\; \mu(y_1 x_1 ... y_{k-1} x_{k-1} y_k \underline{x}_k)$$

With Bayes' Rule, the probability of input $x_1 ... x_k$ if system outputs $y_1 ... y_k$ is

$$\mu(y_1 \underline{x}_1 ... y_k \underline{x}_k) \;=\; \mu(y\!\underline{x}_1) \cdot \mu(y\!x_1 y\!\underline{x}_2) \cdot \;...\; \cdot \mu(y\!x_{<k} y\!\underline{x}_k)$$

Underlined arguments are probability variables, all others are conditions.

μ is called a chronological probability distribution.

Iterative AIμ Model

The Expectimax sequence/algorithm: Take reward expectation over the x_i and maximum over the y_i in chronological order to incorporate correct dependency of x_i and y_i on the history.

$$V_{km}^{best}(\dot{y}\dot{x}_{<k}) \;=\; \max_{y_k}\sum_{x_k}\max_{y_{k+1}}\sum_{x_{k+1}} \;...\; \max_{y_m}\sum_{x_m} (r(x_k)+...+r(x_m))\cdot\mu(\dot{y}\dot{x}_{<k}\underline{yx}_{k:m})$$

$$\dot{y}_k \;=\; \text{maxarg}_{y_k}\sum_{x_k}\max_{y_{k+1}}\sum_{x_{k+1}} \;...\; \max_{y_{m_k}}\sum_{x_{m_k}} (r(x_k)+...+r(x_{m_k}))\cdot\mu(\dot{y}\dot{x}_{<k}\underline{yx}_{k:m_k})$$

This is the essence of Decision Theory.

Decision Theory = Probability + Utility Theory

Functional AIμ $\equiv$ Iterative AIμ

The functional and iterative AIμ models behave identically with the natural identification

$$\mu(y\underline{x}_{1:k}) \;=\; \sum_{q:q(y_{1:k})=x_{1:k}} \mu(q)$$

Remaining Problems:

- Computational aspects.
- The true prior probability is usually not (even approximately not) known.

Limits we are interested in

$$
\begin{array}{ccccccccc}
1 & \ll & \langle l(y_k x_k)\rangle & \ll & k & \ll & T & \ll & |Y\times X| \\
1 & \stackrel{a}{\ll} & 2^{16} & \stackrel{b}{\ll} & 2^{24} & \stackrel{c}{\ll} & 2^{32} & \stackrel{d}{\ll} & 2^{65536}
\end{array}
$$

(a) The agents interface is wide.

(b) The interface can be sufficiently explored.

(c) The death is far away.

(d) Most input/outputs do not occur.

These limits are never used in proofs but ...

... we are only interested in theorems which do not degenerate under the above limits.

Induction = Predicting the Future

Extrapolate past observations to the future, but how can we know something about the future?

Philosophical Dilemma:

- Hume's negation of Induction.
- Epicurus' principle of multiple explanations.
- Ockhams' razor (simplicity) principle.
- Bayes' rule for conditional probabilities.

Given sequence $x_1...x_{k-1}$ what is the next letter x_k?

If the true prior probability $\mu(\underline{x}_1...\underline{x}_n)$ is known, and we want to minimize the number of prediction errors, then the optimal scheme is to predict the x_k with highest conditional μ probability, i.e. $\mathrm{maxarg}_{y_k}\,\mu(x_{<k}\underline{y}_k)$.

Solomonoff solved the problem of unknown prior μ by introducing a universal probability distribution ξ related to Kolmogorov Complexity.

Algorithmic Complexity Theory

The Kolmogorov Complexity of a string x is the length of the shortest (prefix) program producing x.

$$K(x) \;:=\; \min_p\{l(p) : U(p) = x\} \quad , \quad U = \text{univ.TM}$$

The universal semimeasure is the probability that output of U starts with x when the input is provided with fair coin flips

$$\xi(\underline{x}) \;:=\; \sum_{p\,:\,U(p)=x*} 2^{-l(p)} \qquad \text{[Solomonoff 64]}$$

Universality property of ξ: ξ dominates every computable probability distribution

$$\xi(\underline{x}) \;\stackrel{\times}{\geq}\; 2^{-K(\rho)}\cdot\rho(\underline{x}) \quad \forall\rho$$

Furthermore, the μ expected squared distance sum between ξ and μ is finite for computable μ

$$\sum_{k=1}^\infty \sum_{x_{1:k}} \mu(\underline{x}_{<k})(\xi(x_{<k}\underline{x}_k) - \mu(x_{<k}\underline{x}_k))^2 \;\stackrel{+}{<}\; \ln 2\cdot K(\mu)$$

[Solomonoff 78] for binary alphabet

Universal Sequence Prediction

$\Rightarrow \xi(x_{<n}\underline{x}_n) \stackrel{n\to\infty}{\longrightarrow} \mu(x_{<n}\underline{x}_n)$ with μ probability 1.

$\Rightarrow$ Replacing μ by ξ might not introduce many additional prediction errors.

General scheme: Predict x_k with prob. $\rho(x_{<k}\underline{x}_k)$.

This includes deterministic predictors as well.

Number of expected prediction errors:
$E_{n\rho} \;:=\; \sum_{k=1}^n \sum_{x_{1:k}} \mu(\underline{x}_{1:k})(1-\rho(x_{<k}\underline{x}_k))$

Θ_ξ predicts x_k of maximal $\xi(x_{<k}\underline{x}_k)$.

$E_{n\Theta_\xi}/E_{n\rho} \;\leq\; 1+O(E_{n\rho}^{-1/2}) \stackrel{n\to\infty}{\longrightarrow} 1$ [Hutter 99]

Θ_ξ is asymptotically optimal with rapid convergence.

For every (passive) game of chance for which there exists a winning strategy, you can make money by using Θ_ξ even if you don't know the underlying probabilistic process/algorithm.

Θ_ξ finds and exploits every regularity.

Definition of the Universal AIξ Model

Universal AI = Universal Induction + Decision Theory

Replace μ^{AI} in decision theory model AIμ by an appropriate generalization of ξ .

$$\xi(\underline{yx}_{1:k}) \;:=\; \sum_{q:q(y_{1:k})=x_{1:k}} 2^{-l(q)}$$

$$\dot y_k \;=\; \max\arg_{y_k} \sum_{x_k} \max_{y_{k+1}} \sum_{x_{k+1}} \ldots \max_{y_{m_k}} \sum_{x_{m_k}} (r(x_k)+\ldots+r(x_{m_k}))\cdot\xi(\dot y\dot x_{<k}\underline{yx}_{k:m_k})$$

Functional form: $\mu(q) \hookrightarrow \xi(q) := 2^{-l(q)}$.

Bold Claim: AIξ is the most intelligent environmental independent agent possible.

Universality of ξ^{AI}

The proof is analog as for sequence prediction. Inputs y_k are pure spectators.

$$\xi(y\!\underline{x}_{1:n}) \;\stackrel{\times}{\geq}\; 2^{-K(\rho)}\rho(y\!\underline{x}_{1:n}) \quad \forall \text{ chronological } \rho$$

Convergence of ξ^{AI} to μ^{AI}

y_i are again pure spectators. To generalize SP case to arbitrary alphabet we need

$$\sum_{i=1}^{|X|}(y_i - z_i)^2 \;\leq \sum_{i=1}^{|X|} y_i \ln\frac{y_i}{z_i} \quad \text{for} \sum_{i=1}^{|X|} y_i = 1 \geq \sum_{i=1}^{|X|} z_i$$

$$\Rightarrow \xi^{AI}(y\!x_{<n}y\!\underline{x}_n) \stackrel{n\to\infty}{\longrightarrow} \mu^{AI}(y\!x_{<n}y\!\underline{x}_n) \text{ with } \mu \text{ prob. 1.}$$

Does replacing μ^{AI} with ξ^{AI} lead to AIξ system with asymptotically optimal behaviour with rapid convergence?

This looks promising from the analogy with SP!

Value Bounds (Optimality of AIξ)

Naive reward bound analogously to error bound for SP

$$V_{1T}^{\mu}(p^*) \;\geq\; V_{1T}^{\mu}(p) - o(...) \quad \forall\mu,p$$

Problem class (set of environments) $\{\mu_0,\mu_1\}$ with $Y=V=\{0,1\}$ and $r_k=\delta_{iy_1}$ in environment μ_i violates reward bound. The first output y_1 decides whether all future $r_k=1$ or 0.

Alternative: μ probability $D_{n\mu\xi}/n$ of suboptimal outputs of AIξ different from AIμ in the first n cycles tends to zero

$$D_{n\mu\xi}/n \to 0 \quad , \quad D_{n\mu\xi} := \langle \sum_{k=1}^{n} 1-\delta_{\dot y_k^\mu,\dot y_k^\xi} \rangle_\mu$$

This is a weak asymptotic convergence claim.

Value Bounds (Optimality of AIξ)

Let $V=\{0,1\}$ and $|Y|$ be large. Consider all (deterministic) environments in which a single complex output y^* is correct ($r=1$) and all others are wrong ($r=0$). The problem class is

$$\{\mu : \mu(y\!\!x_{<k}y_k\underline{1}) = \delta_{y_k y^*},\ K(y^*) = _\lfloor \log_2 |Y|_\rfloor\}$$

Problem: $D_{k\mu\xi} \leq 2^{K(\mu)}$ is the best possible error bound we can expect, which depends on $K(\mu)$ only. It is useless for $k \ll |Y| \stackrel{\times}{=} 2^{K(\mu)}$, although asymptotic convergence satisfied.

But: A bound like $2^{K(\mu)}$ reduces to $2^{K(\mu|\dot{x}_{<k})}$ after k cycles, which is $O(1)$ if enough information about μ is contained in $\dot{x}_{<k}$ in any form.

Separability Concepts

... which might be useful for proving reward bounds

- Forgetful μ.
- Relevant μ.
- Asymptotically learnable μ.
- Farsighted μ.
- Uniform μ.
- (Generalized) Markovian μ.
- Factorizable μ.
- (Pseudo) passive μ.

Other concepts

- Deterministic μ.
- Chronological μ.

Sequence Prediction (SP)

Sequence $z_1z_2z_3...$ with true prior $\mu^{SP}(z_1z_2z_3...)$.

AIμ Model:

$y_k \quad =$ prediction for z_k.

$r_{k+1} = \delta_{y_k z_k} = 1/0$ if prediction was correct/wrong.

$x'_{k+1} = \epsilon$.

$$\mu^{AI}(y_1\underline{r}_1...y_k\underline{r}_k) = \mu^{SP}(\underline{\delta_{y_1r_1}...\delta_{y_kr_k}}) = \mu^{SP}(\underline{z_1...z_k})$$

Choose horizon h_k arbitrary $\Rightarrow$

$$\dot{y}_k^{AI\mu} \;=\; \underset{y_k}{\text{maxarg}}\,\mu(\dot{z}_1...\dot{z}_{k-1}\underline{y}_k) \;=\; \dot{y}_k^{SP\Theta_\mu}$$

AIμ always reduces exactly to XXμ model if XXμ is optimal solution in domain XX.

AIξ model differs from SPΘ_ξ model. For $h_k\!=\!1$

$$\dot{y}_k^{AI\xi} \;=\; \underset{y_k}{\text{maxarg}}\,\xi(\dot{y}\dot{r}_{<k}y_k\underline{1}) \;\neq\; \dot{y}_k^{SP\Theta_\xi}$$

Weak error bound: $E_{n\xi}^{AI} \;\overset{\times}{<}\; 2^{K(\mu)} \;<\; \infty$ for deterministic μ.

Strategic Games (SG)

- Consider strictly competitive strategic games like chess.
- Minimax is best strategy if both Players are rational with unlimited capabilities.
- Assume that the environment is a minimax player of some game $\Rightarrow \mu^{AI}$ uniquely determined.
- Inserting μ^{AI} into definition of $\dot{y}_k^{AI}$ of AIμ model reduces the expecimax sequence to the minimax strategy ($\dot{y}_k^{AI} = \dot{y}_k^{SG}$).
- As $\xi^{AI} \rightarrow \mu^{AI}$ we expect AIξ to learn the minimax strategy for any game and minimax opponent.
- If there is only non-trivial reward $r_k \in \{win, loss, draw\}$ at the end of the game, repeated game playing is necessary to learn from this very limited feedback.
- AIξ can exploit limited capabilities of the opponent.

Function Minimization (FM)

Approximately minimize (unknown) functions with as few function calls as possible.

Applications:

- Traveling Salesman Problem (bad example).
- Minimizing production costs.
- Find new materials with certain properties.
- Draw paintings which somebody likes.

$$\mu^{FM}(y_1\underline{z}_1...y_n\underline{z}_n) \;:=\; \sum_{f:f(y_i)=z_i\;\forall 1\leq i\leq n} \mu(f)$$

Trying to find y_k which minimizes f in the next cycle does not work.

General Ansatz for FMμ/ξ:

$$\dot{y}_k = \operatorname*{minarg}_{y_k}\sum_{z_k}...\min_{y_T}\sum_{z_T}(\alpha_1 z_1 + ... + \alpha_T z_T)\cdot\mu(\dot{y}\dot{z}_1...y\underline{z}_T)$$

Under certain weak conditions on α_i, f can be learned with AIξ.

Supervised Learning by Examples (EX)

Learn functions by presenting $(z,f(z))$ pairs and ask for function values of z' by presenting $(z',?)$ pairs.

More generally: Learn relations $R\ni(z,v)$.

Supervised learning is much faster than reinforcement learning.

The AIμ/ξ model:

$x'_k \;\; = (z_k,v_k) \in R\cup(Z\times\{?\}) \subset Z\times(Y\cup\{?\}) = X'$

$y_{k+1} =$ guess for true v_k if actual $v_k = ?$.

$r_{k+1} = 1$ iff $(z_k,y_{k+1})\in R$

AIμ is optimal by construction.

Supervised Learning by Examples (EX)

The AIξ model:

- Inputs x'_k contain much more than 1 bit feedback per cycle.
- Short codes dominate ξ
- The shortest code of examples (z_k, v_k) is a coding of R and the indices of the (z_k, v_k) in R.
- This coding of R evolves independently of the rewards r_k.
- The system has to learn to output y_{k+1} with $(z_k, y_{k+1}) \in R$.
- As R is already coded in q, an additional algorithm of length $O(1)$ needs only to be learned.
- Credits r_k with information content $O(1)$ are needed for this only.
- AIξ learns to learn supervised.

Computability and Monkeys

SPξ and AIξ are not really uncomputable (as often stated) but ...

$\dot{y}_k^{AI\xi}$ is only asymptotically computable/approximable with slowest possible convergence.

Idea of the typing monkeys:

- Let enough monkeys type on typewriters or computers, eventually one of them will write Shakespeare or an AI program.
- To pick the right monkey by hand is cheating, as then the intelligence of the selector is added.
- Problem: How to (algorithmically) select the right monkey.

The Timebounded AIξ Model

An algorithm p^{best} can be/has been constructed for which the following holds:

Result/Theorem:

- Let p be any (extended) chronological program
- with length $l(p) \leq \tilde{l}$ and computation time per cycle $t(p) \leq \tilde{t}$
- for which there exists a proof of VA(p), i.e. that p is a valid approximation, of length $\leq l_P$.
- Then an algorithm p^{best} can be constructed, depending on $\tilde{l}$,$\tilde{t}$ and l_P but not on knowing p
- which is effectively more or equally intelligent according to $\succeq^c$ than any such p.
- The size of p^{best} is $l(p^{best}) = O(\ln(\tilde{l} \cdot \tilde{t} \cdot l_P))$,
- the setup-time is $t_{setup}(p^{best}) = O(l_P^2 \cdot 2^{l_P})$,
- the computation time per cycle is $t_{cycle}(p^{best}) = O(2^{\tilde{l}} \cdot \tilde{t})$.

Aspects of AI included in AIξ

All known and unknown methods of AI should be directly included in the AIξ model or emergent.

Directly included are:

- Probability theory (probabilistic environment)
- Utility theory (maximizing rewards)
- Decision theory (maximizing expected reward)
- Probabilistic reasoning (probabilistic environment)
- Reinforcement Learning (rewards)
- Algorithmic information theory (universal prob.)
- Planning (expectimax sequence)
- Heuristic search (use ξ instead of μ)
- Game playing (see SG)
- (Problem solving) (maximize reward)
- Knowledge (in short programs q)
- Knowledge engineering (how to train AIξ)
- Language or image processing (has to be learned)

Other Aspects of AI not included in AIξ

Not included: Fuzzy logic, Possibility theory, Dempster-Shafer, ...

Other methods might emerge in the short programs q if we would analyze them.

AIξ seems not to lack any important known methodology of AI, apart from computational aspects.

Outlook & Uncovered Topics

- Derive general and special reward bounds for AIξ.
- Downscale AIξ in more detail and to more problem classes analog to the downscaling of SP to Minimum Description Length and Finite Automata.
- There is no need for implementing extra knowledge, as this can be learned.
- The learning process itself is an important aspect.
- Noise or irrelevant information in the inputs do not disturb the AIξ system.

Conclusions

- We have developed a parameterless model of AI based on Decision Theory and Algorithm Information Theory.
- We have reduced the AI problem to pure computational questions.
- A formal theory of something, even if not computable, is often a great step toward solving a problem and also has merits in its own.
- All other systems seem to make more assumptions about the environment, or it is far from clear that they are optimal.
- Computational questions are very important and are probably difficult. This is the point where AI could get complicated as many AI researchers believe.
- Nice theory yet complicated solution, as in physics.

The Big Questions

- Is non-computational physics relevant to AI? [Penrose]
- Could something like the number of wisdom Ω prevent a simple solution to AI? [Chaitin]
- Do we need to understand consciousness before being able to understand AI or construct AI systems?
- What if we succeed?