Texts in Theoretical Computer Science
An EATCS Series

Marcus Hutter

Universal Artificial Intelligence

Sequential Decisions Based on Algorithmic Probability

Springer

Author

Dr. Marcus Hutter
Istituto Dalle Molle
di Studi sull'Intelligenza
Artificiale (IDSIA)
Galleria 2
CH-6928 Manno-Lugano
Switzerland
marcus@idsia.ch
www.idsia.ch/~marcus

Series Editors

Prof. Dr. Wilfried Brauer
Institut für Informatik der TUM
Boltzmannstr. 3, 85748 Garching, Germany
Brauer@informatik.tu-muenchen.de

Prof. Dr. Grzegorz Rozenberg
Leiden Institute of Advanced Computer Science
University of Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
rozenber@liacs.nl

Prof. Dr. Arto Salomaa
Turku Centre for Computer Science
Lemminkäisenkatu 14 A, 20520 Turku, Finland
asalomaa@utu.fi

Library of Congress Control Number: 2004112980

ACM Computing Classification (1998): I.3, I.2.6, F.0, F.1.3, F.4.1, E.4, G.3

ISBN 3-540-22139-5 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Cover design: KünkelLopka, Heidelberg
Typesetting: by the author
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Printed on acid-free paper 45/3142/YL - 5 4 3 2 1 0

Preface

Personal motivation. The dream of creating artificial devices that reach or outperform human intelligence is an old one. It is also one of the dreams of my youth, which have never left me. What makes this challenge so interesting? A solution would have enormous implications on our society, and there are reasons to believe that the AI problem can be solved in my expected lifetime. So, it's worth sticking to it for a lifetime, even if it takes 30 years or so to reap the benefits.

The AI problem. The science of artificial intelligence (AI) may be defined as the construction of intelligent systems and their analysis. A natural definition of a *system* is anything that has an input and an output stream. Intelligence is more complicated. It can have many faces like creativity, solving problems, pattern recognition, classification, learning, induction, deduction, building analogies, optimization, surviving in an environment, language processing, and knowledge. A formal definition incorporating every aspect of intelligence, however, seems difficult. Most, if not all known facets of intelligence can be formulated as goal driven or, more precisely, as maximizing some utility function. It is, therefore, sufficient to study goal-driven AI; e.g. the (biological) goal of animals and humans is to survive and spread. The goal of AI systems should be to be useful to humans. The problem is that, except for special cases, we know neither the utility function nor the environment in which the agent will operate in advance. The major goal of this book is to develop a theory that solves these problems.

The nature of this book. The book is theoretical in nature. For most parts we assume availability of unlimited computational resources. The first important observation is that this does not make the AI problem trivial. Playing chess optimally or solving NP-complete problems become trivial, but driving a car or surviving in nature do not. This is because it is a challenge itself to well-define the latter problems, not to mention presenting an algorithm. In other words: The AI problem has not yet been well defined. One may view the book as a suggestion and discussion of such a mathematical definition of AI.

Extended abstract. The *goal* of this book is to develop a universal theory of sequential decision making akin to Solomonoff's celebrated universal theory of induction. Solomonoff derived an optimal way of predicting future data, given previous observations, provided the data is sampled from a computable probability distribution. Solomonoff's unique predictor is universal in the sense that it applies to every prediction task and is the output of a universal Turing machine with random input. We extend this approach to derive an optimal rational reinforcement learning agent, called AIXI, embedded in an unknown environment. The *main idea* is to replace the unknown environmental distribution μ in the Bellman equations by a suitably generalized universal distribution ξ. The state space is the space of complete histories. AIXI is a universal theory without adjustable parameters, making no assumptions about the environment except that it is sampled from a computable distribution. From an algorithmic complexity perspective, the AIXI model generalizes optimal passive universal induction to the case of active agents. From a decision-theoretic perspective, AIXI is a suggestion of a new (implicit) "learning" algorithm, which may overcome all (except computational) problems of previous reinforcement learning algorithms.

Chapter 1. We start with a survey of the contents and main results in this book.

Chapter 2. How and in which sense induction is possible at all has been subject to long philosophical controversies. Highlights are Epicurus' principle of multiple explanations, Occam's razor, and Bayes' rule for conditional probabilities. Solomonoff elegantly unified all these aspects into one formal theory of inductive inference based on a universal probability distribution ξ, which is closely related to Kolmogorov complexity $K(x)$, the length of the shortest program computing x. We classify the (non)existence of universal priors for several generalized computability concepts.

Chapter 3. We prove rapid convergence of ξ to the unknown true environmental distribution μ and tight loss bounds for arbitrary bounded loss functions and finite alphabet. We show Pareto optimality of ξ in the sense that there is no other predictor that performs better or equal in all environments and strictly better in at least one. Finally, we give an Occam's razor argument showing that predictors based on ξ are optimal. We apply the results to games of chance and compare them to predictions with expert advice. All together this shows that Solomonoff's induction scheme represents a universal (formal, but incomputable) solution to all *passive* prediction problems.

Chapter 4. Sequential decision theory provides a framework for finding optimal reward-maximizing strategies in *reactive* environments (e.g. chess playing as opposed to weather forecasting), assuming the environmental probability distribution μ is known. We present this theory in a very general form (called AIμ model) in which actions and observations may depend on arbitrary past events. We clarify the connection to the Bellman equations and discuss minor parameters including (the size of) the I/O spaces and the lifetime of the agent and their universal choice which we have in mind. Optimality of AIμ is obvious by construction.

Chapter 5. Reinforcement learning algorithms are usually used in the case of unknown μ. They can succeed if the state space is either small or has ef-

fectively been made small by generalization techniques. The algorithms work only in restricted, (e.g. Markovian) domains, have problems with optimally trading off exploration versus exploitation, have nonoptimal learning rate, are prone to diverge, or are otherwise ad hoc. The formal solution proposed in this book is to generalize the universal prior ξ to include actions as conditions and replace μ by ξ in the AIμ model, resulting in the AIXI model, which we claim to be universally optimal. We investigate what we can expect from a universally optimal agent and clarify the meanings of *universal*, *optimal*, etc. We show that a variant of AIXI is self-optimizing and Pareto optimal.
Chapter 6. We show how a number of AI problem classes fit into the general AIXI model. They include sequence prediction, strategic games, function minimization, and supervised learning. We first formulate each problem class in its natural way for known μ, and then construct a formulation within the AIμ model and show their equivalence. We then consider the consequences of replacing μ by ξ. The main goal is to understand in which sense the problems are solved by AIXI.
Chapter 7. The major drawback of AIXI is that it is incomputable, or more precisely, only asymptotically computable, which makes an implementation impossible. To overcome this problem, we construct a modified model AIXItl, which is still superior to any other time t and length l bounded algorithm. The computation time of AIXItl is of the order $t \cdot 2^l$. A way of overcoming the large multiplicative constant 2^l is presented at the expense of an (unfortunately even larger) additive constant. The constructed algorithm M_p^ε is capable of solving all well-defined problems p as quickly as the fastest algorithm computing a solution to p, save for a factor of $1+\varepsilon$ and lower-order additive terms. The solution requires an implementation of first-order logic, the definition of a universal Turing machine within it and a proof theory system.
Chapter 8. Finally we discuss and remark on some otherwise unmentioned topics of general interest. We also critically review what has been achieved in this book, including assumptions, problems, limitations, performance, and generality of AIXI in comparison to other approaches to AI. We conclude the book with some less technical remarks on various philosophical issues.

Prerequisites. I have tried to make the book as self-contained as possible. In particular, I provide all necessary background knowledge on algorithmic information theory in Chapter 2 and sequential decision theory in Chapter 4. Nevertheless, some prior knowledge in these areas could be of some help. The chapters have been designed to be readable independently of one another (after having read Chapter 1). This necessarily implies minor repetitions. Additional information to the book (FAQs, errata, prizes, ...) is available at http://www.idsia.ch/~marcus/ai/uaibook.htm.

Problem classification. Problems are included at the end of each chapter of different motivation and difficulty. We use Knuth's rating scheme for exercises [Knu73] in slightly adapted form (applicable if the material in the corresponding chapter has been understood). In-between values are possible.

C00 *Very easy.* Solvable from the top of your head.
C10 *Easy.* Needs 15 minutes to think, possibly pencil and paper.
C20 *Average.* May take 1–2 hours to answer completely.
C30 *Moderately difficult or lengthy.* May take several hours to a day.
C40 *Quite difficult or lengthy.* Often a significant research result.
C50 *Open research problem.* An obtained solution should be published.

The rating is possibly supplemented by the following qualifier(s):

i Especially *interesting* or *instructive* problem.
m Requires more or higher *math* than used or developed here.
o *Open* problem; could be worth publishing; see web for prizes.
s *Solved* problem with published solution.
u *Unpublished* result by the author.

The problems represent an important part of this book. They have been placed at the end of each chapter in order to keep the main text better focused.

Acknowledgements. I would like to thank all those people who in one way or another have contributed to the success of this book. For interesting discussions I am indebted to Jürgen Schmidhuber, Ray Solomonoff, Paul Vitányi, Peter van Emde Boas, Richard Sutton, Leslie Kaelbling, Leonid Levin, Peter Gács, Wilfried Brauer, and many others. Shane Legg, Jan Poland, Viktor Zhumatiy, Alexey Chernov, Douglas Eck, Ivo Kwee, Philippa Hutter, Paul Vitányi, and Jürgen Schmidhuber gave valuable feedback on drafts of the book. Thanks also collectively to all other IDSIAnies and to the Springer team for the pleasant working atmosphere and their support. This book would not have been possible without the financial support of the SNF (grant no. 2000-61847.00). Thanks also to my father, who taught me to think sharply and to my mother who taught me to do what one enjoys. Finally, I would like to thank my wife and children who patiently supported my decision to write this book.

Lugano, Switzerland, August 2004 *Marcus Hutter*

Contents

Tables, Figures, Theorems, ...

Notation

The following is a list of commonly used notation. The first entry is the symbol itself, followed by its meaning or name (if any) and the page number where the definition appears. Some standard symbols like $\mathbb{R}$ are not defined in the text. There appears a * in place of the page number for these symbols.

Symbol	**Explanation**	**Page**
[C35s]	classification of problems	viii
[Hut04b]	paper, book or other reference	*
(5.3)	label/reference for a formula/theorem/definition/...	*
∞	infinity	*
$\{a,...,z\}$	set containing elements $a,b,...,y,z$. $\{\}$ is the empty set	*
$[a,b)$	interval on the real line, closed at a and open at b	*
$\cap,\cup,\setminus,\in$	set intersection, union, difference, membership	*
$\wedge,\vee,\neg$	Boolean conjunction (and), disjunction (or), negation (not)	*
$\subseteq,\subset$	subset, proper subset	*
$\Rightarrow$	implies	*
$\Leftrightarrow$	equivalence, if and only if, iff	*
$\Box$	q.e.d. (Latin), which was to be demonstrated	*
$\forall,\exists$	for all, there exists	*
$\approx,\lesssim,\gtrsim$	approximately equal, less equal, greater equal	33
$\ll,\gg$	much smaller/greater than	*
$\equiv$	equivalent, identical, equal by definition	*
$\cong$	isomorphic	*
$:=$	define as	*
$\widehat{=}$	corresponds to, informal equality	*
$\sim$	asymptotically proportional to	33
$\propto$	proportional to	*
$=,\neq$	equal to, not equal to	*
$+,-,\cdot,/$	standard arithmetic operations: sum, difference, product, ratio	*
$\sqrt{\ }$	square root	*
$\leq,\geq,<,>$	standard inequalities	*
$\|\mathcal{S}\|,\|a\|$	size/cardinality of set $\mathcal{S}$, absolute value of a	*

$\to$	mapping, approaches, Boolean implication	*
$\to$	converge to each other	33
$\lim\limits_{n\to\infty}$	limiting value of argument for n tending to infinity	*
$\leadsto$	replace with	*
$\lceil x\rceil$	ceiling of x: smallest integer larger or equal than x	*
$\lfloor x\rfloor$	floor of x: largest integer smaller or equal than x	*
δ_{ab}	Kronecker symbol, $\delta_{ab}=1$ if $a=b$ and 0 otherwise	*
$\sum_{k=1}^n$	summation from $k=1$ to n	*
$\sum_x'$	summation over x for which $\mu(x)\neq 0$	69
$\prod_{k=1}^n$	product from $k=1$ to n	*
$\int,\int_a^b dx$	Lebesgue integral, integral from a to b over x	*
i,k,n,t	natural numbers	33
x,y,z	finite strings	33
min/max	min-/maximal element of set: $\min_{x\in\mathcal{X}} f(x)=\min\{f(x):x\in\mathcal{X}\}$	*
argmin	$\text{argmin}_x f(x)$ is the x minimizing $f(x)$ (ties broken arbitrarily)	*
l.h.s.	left-hand side	*
r.h.s.	right-hand side	*
w.r.t.	with respect to	*
e.g.	exempli gratia (Latin), for example	*
i.e.	id est (Latin), that is	*
etc.	et cetera (Latin), and so forth	*
cf.	confer (Latin, imperative of conferre), compare with	*
et al.	et alii (Latin), and others	*
q.e.d.	quod erat demonstrandum (Latin), which was to be shown	*
i.i.d.	independent identically distributed (random variables)	*
iff	if and only if	*
w.p.1/i.p.	with probability 1 / in probability	71
i.m./i.m.s.	in the mean / in mean sum	71
log	logarithm to some basis	*
$\log_b$	logarithm to basis b	*
ln	natural logarithm to basis e=2.71828...	*
e	base of natural logarithm e=2.71828...	*
$\mathbb{R}$	set of real numbers	*
$\mathbb{R}^+$	set of nonnegative real numbers	*
$\mathbb{N}$	set of natural numbers $\{1,2,3,...\}$	33
$\mathbb{N}_0$	set of natural numbers including zero $\{0,1,2,3,...\}$	33
$\mathbb{Z}$	set of integers $\{...,-2,-1,0,1,2,3,...\}$	*
$\mathbb{Q}$	set of rational numbers $\{\frac{n}{d}\}$	*

$\mathbb{B}=\{0,1\}$	binary alphabet	*
$y_t \in \mathcal{Y}$	action (output of agent) in cycle t, followed by ...	128
$x_t \in \mathcal{X}$	perception (feedback/input to agent) in cycle t	45, 128
$o_t \in \mathcal{O}$	informative input/observation in cycle t	128
$r_t \in \mathcal{R} \subset \mathbb{R}$	reward in cycle t	128
ε	some small positive real number	*
ϵ	empty string	33
$*$	wildcard for some string (prefix, finite, or infinite)	33
$x_{1:n}$	$=x_1...x_n=$ string of length n	45, 68, 128
$x_{<t}$	$=x_1...x_{t-1}=$ string of length $t-1$	45, 68, 128
$yx_{k:n}$	action-perception sequence $y_k x_k ... y_n x_n$	128
$\dot{y}\dot{x}_{<k}$	actually realized action-perception sequence $\dot{y}_1\dot{x}_1...\dot{y}_{k-1}\dot{x}_{k-1}$	130
ω	infinite sequence, elementary event	33
Ω	sample space	42, 68
$\Gamma_{x_{1:n}}$	$=\{\omega:\omega_{1:n}=x_{1:n}\}=$ cylinder set	46, 68
$\ell(x)$	length of string x	33
$\langle o \rangle$	coding of object o	33
$\langle x,y \rangle$	uniquely decodable pairing of x and y	33
x'	prefix coding of x	33
$O(),o()$	big and small oh-notation	33
$a \overset{+}{\leq} b$	less within an additive const., i.e. $a \leq b+O(1)$. Similarly $\overset{+}{=}$	33
$a \overset{\times}{\leq} b$	less within a multiplicative const., i.e. $a=O(b)$. Similarly $\overset{\times}{=}$	33
$K(x)$	prefix Kolmogorov complexity of string x	37
$Km(x_{1:n})$	monotone (Kolmogorov) complexity of string $x_{1:n}$	47, 190
$K(o_1 \| o_2)$	Kolmogorov complexity of object o_1, given object o_2	37
$M \overset{\times}{=} \xi_U$	Solomonoff-Levin's universal semimeasure	46, 48
$\mathcal{M}=\{\nu\}$	(usually countable) set of (semi)measures	48, 81
EC	$\in$ {AI, SP, FM, EX, SG, ...} is an environmental class	*
AI	artificial or algorithmic intelligence,	2
	most general computational environmental class	130, 154
SP	sequence prediction	187
CF	classification	108
SG	strategic two-player informed zero-sum games	192
FM	function minimization	197
EX	supervised learning (by examples)	204
pd	probability density function / distribution / measure	*
$\rho(x_{1:n})$	probability of string/sequence starting with $x_{1:n}$	46, 68
$\mu \in \mathcal{M}$	true generating environmental pd	68

$\mathbf{E}$	expectation value, usually w.r.t. the true distribution μ	68
$\mathbf{P}$	probability, usually w.r.t. the true distribution μ	68
$\mu(x_1\underline{x}_2x_3\underline{x}_4)$	μ probability that the 2^{nd} and 4^{th} symbols of a string are x_2 and x_4, given the 1^{st} and 3^{rd} symbols are x_1 and x_3	132
$\nu\in\mathcal{M}$	any pd in $\mathcal{M}$	70
ρ	any pd not necessarily in $\mathcal{M}$ usually specifying a policy	68
ξ	$=\sum_{\nu\in\mathcal{M}}w_\nu\nu=$ mixture (belief) pd	48, 70
w_ν	prior degree of belief in ν –or– weight of ν	48, 70
ρ^{EC}	pd of environmental argument type EC	185
ξ^{EC}	mixture distribution of type EC for class EC	185
$\ell_{x_ty_t}$	incurred loss when predicting y_t and x_t is next symbol	86
$l_{t\nu}^\Lambda$	ν-expected instantaneous loss in step t of predictor Λ	99, 87
$L_{n\nu}^\Lambda$	ν-expected cumulative loss of steps $1...n$ of predictor Λ	100
Θ_ρ	predictor with minimal number of ρ-expected errors	82
Λ_ρ	predictor that minimizes the ρ-expected loss	87
$e_{t\nu}^\Theta$	ν-probability that Θ-predictor errs in step t	83
$E_{n\nu}^\Theta$	ν-expected number of errors in steps $1...n$ of predictor Θ	83
$L_n^\Lambda\equiv L_{n\mu}^\Lambda$	abbreviation for true μ-expected loss	86
$V_{km}^{p\nu}(\dot{y}\dot{x}_{<k})$	value of policy p in environment ν given history $\dot{y}\dot{x}_{<k}$	153
y_t^Λ	prediction/decision/action of predictor Λ in step t	87
y_k^p	action of policy p in cycle k	*
γ_k	discounting sequence	159
Γ_k	value function normalization $(\sum_{i=k}^\infty\gamma_k)$	159
m,h	agent's lifespan, horizon	129, 169
p	agent's policy	126
q	deterministic environment	126
p^ν	policy that maximizes value V_ν^p	130
$V_\mu^*\equiv V_{1m}^{p^\mu\mu}$	true or generating value	130
$V_\xi^*\equiv V_{1m}^{p^\xi\xi}$	universal value	146
$D_n\equiv D_{n\mu}^\xi$	relative entropy between μ and ξ for the first n cycles	73