Index

\(E^3\) 121
\(\alpha\) 30
\(\alpha\)-functions 374
\(\gamma\) 14

\(\alpha\beta\)-pruning 551
absorbing state 13, 121
abstract action
  sequential 316
abstract actions 297
abstraction
  in Poker 564
action 10
  applicable 10
  precondition 11
action language 256
action-selection networks 329
actor critic 212, 230, 234
ADHDP 236
Cacla 236, 238
  natural actor critic 235, 238
actor only 230
actor-critic algorithm 32
Actor-critic methods 369
agent 5
agnostic 189
alternating maximization 481
approximate linear programming
  in Tetris 557
approximate policy iteration 78
  optimistic 89
arms race 342
artificial general intelligence 293
asymptotic improvement 153
asynchronous updating 25
automatic feature construction 563
average loss 182
Backgammon 542
  Neurogammon 543
  TD-Gammon 543
backpropagation 543, 546
  Tangent-Prop 546
backup operator 21, 24
Baldwin effect 335
ball-in-a-cup 599
basal ganglia 521
basis function 81
batch learning 45
  algorithms 52
  applications 69
  problem 46
Bayesian actor-critic 369
Bayesian DP 127
Bayesian game 486
Bayesian Multi-Task Reinforcement Learning 377
Bayesian policy gradient 368
Bayesian Q-learning 361
Bayesian quadrature 366
Bayesian reinforcement learning 359
BEB 128
BEETLE 127, 376
behavioral cloning 36
belief 392
Belief monitoring 373
belief state MDP 127
belief update 392
Bellman backup 374
Bellman equation 16, 77, 373
projected, see projected Bellman equation
Bellman optimality equation 16
Bellman residual minimization 80, 84
best response policy 568
Bgblitz 543
bias-variance tradeoffs 381
binary action search 132
Blocks World 258
bootstrapping 30
BOSS 127
bounded policy iteration 496
branching factor 542
Cacla 236, 238
cart pole
double-pole cart pole 238
cart-pole balancing 133
case-based learning 570
in real-time strategy games 560
cellular encodings 343
Checkers 548, 564
Chess 545
knight fork 549
Morph 564
NeuroChess 546
tutoring 566
UCT 555
classical conditioning 508
CMA-ES 221, 238, 330
in Tetris 557
cognitive architecture 257
communication 497
competing conventions problem 332
competitions
AIIDE Starcraft Competition 561
Ms. Pac-Man Competition 562
competitive fitness sharing 342
complexity of dynamic programming 24
crash assignment 7
cross-entropy method
in Tetris 557
cross-entropy optimization 221, 493
curse of dimensionality 582
Darwinian evolution 334
DBN-E 3 124
Dec-POMDP 471
brute force search 480
decision rule 477
dynamic programming 489
factored 494
history 475
infinite horizon 495
model 473
optimal Q-value function 484
policy 476
TOI 494
transition independent 494
Dec-POMDP-Com 497
decentralized POMDP 471
decision-theoretic planning 18
decision-theoretic regression 262
Deep Blue 545
Deep Fitted Q Iteration 67
delayed feedback 7
density estimation 330
DFQ 67
difference functions 340
difficulty scaling 569
direct policy search 212, 234, 556
dynamic scripting 559
partial observability 567
Dirichlet distribution 372
Dirichlet mixture 379
Dirichlet process 380
discount factor 14
diversity of gameplay 569
domain knowledge 360, 553
in games 561
dopamine 518
DRE 132
DYNA 38
Dyna 118
Dyna-2 551
Dyna-2 119
Dynamic Bayesian Network 122
Dynamic Programming 373
dynamic programming 19
  Dec-POMDP 489
  memory-bounded 492
  point-based 491
  dynamic scripting 570
    difficulty scaling 569
    in real-time strategy games 559
  EDI-CR 494
  eligibility trace 37
  eligibility traces 227
    performance in Backgammon 545
  Elo rating 545
    of patterns 551
  envelope 26
  environment 5
  episodic task 12
  ESP 341
  estimation of distribution algorithms 330
  evaluative feedback 8
  event-driven Dec-MDP 494
  evolutionary algorithms 221
  evolutionary computation 325
    coevolution 339
    neuroevolution 327
    on-line 346
    steady-state 338
  evolutionary function approximation 335
  evolutionary game theory 550
  evolutionary learning
    in Starcraft 560
    in Tetris 557
    in Wargus 560
  evolutionary strategies 221, 234
    CMA-ES 221, 238
    natural evolutionary strategies 221, 234
  EvolutionChamber 560
  experience replay 50
  experiment
    double-pole cart pole 238
  explicit fitness sharing 334
  exploration 28, 34, 88, 126
    automatic 544
    Boltzmann 29
    Boltzmann exploration 232
    ε-greedy 565
    Gaussian exploration 232
    in games 564
    softmax 29
  exploration gain 377
  exploration overhead 49
  exploration-exploitation tradeoff 8
  exploration/exploitation tradeoff 360, 373
  factored Dec-POMDP 494
  factored state 122
  factored-r-max 124
  family of relational MDPs 260
  feature transfer 157, 161, 162
  finite horizon task 12
  Finite Sample Analysis 380
  first-order (logical) basis functions 266
  first-order decision-theoretic regression 264
  Fisher kernels 368
  fitness function 325
  Fitted Q Iteration 55
  Fitted-R-Max 132
  fitting 50
  fog of war 566
  forward-sweep policy computation 485
  FQI 55
  fun gameplay 568
    balanced strategies 571
  function approximation 34, 208, 212, 592
    discretization 214
    fuzzy sets 217
    in Tetris 556
    linear 213
    non-linear 217
    tile coding 214
  game balance 571
  games 539
    Atari 2600 games 562
    Backgammon 542
    Baldur’s Gate 569
    Black & White 570
    Bridge 567
    Checkers 564
    Chess 545
    Civilization 570
    Creatures 570
    Diplomacy 564
    general gameplay 564
    Go 550
    Hearts 564
    Knock’em 569
    Last Night on Earth 571
    Magic: The Gathering 571
Ms. Pac-Man 564
NERO 570
Neverwinter Nights 570, 571
Othello 564
Poker 564, 568
Pong 569
real-time strategy 558
Suicide chess 550
SZ-Tetris 556
teaching games 570
Tetris 555
Gaussian process 363
Gaussian Process Temporal Difference 363
generalization 114
generalized multi-agent A* 488
generalized policy iteration 18
generative and developmental systems 327, 343
generative model 113
genetic algorithms 325, 493
GMAA* 488
GNUbg 543
GNUchess 547
Go 550
nakade 555
tutoring 566
goal specification 585
goal state 12
gradient descent 219
gradient free optimization 220
gradient temporal-difference learning 226
growing batch learning
problem 48
guidance 36
heuristics 26
equivalent experience 553
RAVE 553
virtual wins/losses 553
HEXQ
exit 315
hierarchical reinforcement learning (HRL) 295
HAMQ 307
HEXQ 315
MAXQ 309
options 306
host/parasite model 342
hyperparameters 373
imitation 594
incremental policy generation 491
indefinite horizon task 12
independent Q-learners 496
indirect encodings 343
inductive logic programming 256
infinite horizon task 12
initial state distribution 12
innovation numbers 333
instance transfer 150
instructive feedback 8
intensantional dynamic programming 262
international planning competition 267
inverse reinforcement learning 585
iterated elimination of dominated policies 491
JESP 481
joint action 445
joint belief 482, 498
jumpstart improvement 153
KADP 52
kernel-based approximate dynamic programming 52
KnightCap 547
known state–action MDP 184, 187, 192, 198
empirical known state–action MDP 184
KWIK 121, 189
Knows What It Knows 189
KWIK-learnable 190
KWIK-learnable MDPs 192
\(\lambda\)-policy iteration
in Tetris 557
L-systems 343
Lamarckian evolution 334
Law of Effect 514
learning classifier systems 335, 336
anticipatory 346
Michigan-style 337
Pittsburgh-style 337
learning rate 30
learning speed improvement 151
least-squares policy evaluation 84
least-squares policy iteration 57, 86
in Tetris 557
online 90
least-squares temporal difference 84
Index

lifted first-order reasoning 266
logical generalization 257
logical language 258
logical matching 263
logical regression 263
lookahead search
  multi-step 547
LSE-R-Max 124
LSPI 57, 132

MAA* 488
Markov 11
Markov decision process 10, 12
Markov game
  definition 455
Markov property 456
maximum likelihood estimation 368
MBBE 128
MBIE 128
memory-bounded dynamic programming
  492
Met-R-Max 124
micromanagement 570
minimax 342
minimax search 547, 555
minimax-optimal strategy 568
missing information
  in games 566
model 12
  learning of 33, 38
model approximation 211
model errors 584
model-based algorithm 17, 19
model-based control 516
Model-Based Reinforcement Learning 372
model-free algorithm 17, 27
model-free control 516
Model-free RL 361
modified policy iteration 25
monomial basis 376
Monte Carlo sampling 33
Monte Carlo Tree Search 115
Monte-Carlo 366
  hidden information 567
  value estimation 552
Monte-Carlo tree search 552
  in Chess 549, 555
  in Go 553
Morph 549, 564
multi-agent A* 488
multi-agent belief 490, 491
multi-agent MDP 494
multi-agent reinforcement learning 496
Multi-task learning 377
myopic policy 377
Nash equilibrium 481, 494
Nash-equilibrium 568
natural actor critic 235, 238
natural evolutionary strategies 221, 234
natural gradient 220
natural policy gradient 233
ND-POMDP 495
NEAT 333
Neural Fitted Q Iteration 61
neural network 543, 546
  as automatic feature construction 563
  as predictive model 546
  for pattern evaluation 551
neural networks 327
  feed-forward 329
  recurrent 329
NeuroChess 546
Neurogammon 543
neurons 340
Neverwinter Nights 570, 571
NFQ 61
niching 332
objective functions 224
observation model 390
off-policy learning 32
offline learning 7
on-policy learning 31
Online algorithms 376
online learning 7
operant conditioning 513
opponent modelling 567
optimal
  hierarchically greedy 304
  hierarchically optimal 304
  recursively optimal 304
optimism in the face of uncertainty 181
  186, 197
optimistic value initialization 29
option transfer 156, 161
ORTS 559
ORTS Competition 561
Othello 564

PAC-Bayesian approach 382
PAC-MDP 179, 382
parallel tasks 558
parameter transfer 151
parameterized action 257
Pareto optimal 494
Pareto optimality 342
parti-game 131
partial observability 388
partial visibility 566
partially observable Markov decision process 11, 372, 390
partially observable stochastic game 493
patterns 551, 554
local 549, 551
move patterns 549
team of 551
perceptual aliasing 388
Perseus 376
PIAGeT 254
PILCO 131
planning 6, 18, 115
player ranking
  Elo rating 545
  rollout analysis 543
tournament 543
playtesting 571
point-based dynamic programming
  Dec-POMDP 491
point-based value iteration 376
Poker 564, 568
  state aggregation 564
policy 13
  ε-greedy 29
  application of 13
deterministic 13
greedy 17, 21
  optimal 13, 16
stochastic 13
policy approximation 212, 229
policy evaluation 18, 20, 77
  least-squares, see least-squares policy evaluation
  projected, see projected policy evaluation
Policy gradient 365
policy gradient 230, 366
natural policy gradient 233
policy gradient theorem 370
policy improvement 18, 21, 78
policy iteration 20, 77
  approximate, see approximate policy iteration
  least-squares, see least-squares policy iteration
policy search 588
POMDP 390
  Bellman equation 394
  learning internal memory 405
  model-based techniques 395
  model-free techniques 404
  point-based methods 401
  value function 394
  value iteration 398
POMDPs
  deep-memory 341
Pong 569
POSG 493
posterior Gaussian process 364
preference function 556, 557
Prior Knowledge 379
prior knowledge 594
prioritized sweeping 33, 38
probabilistic inference for planning 280
probabilistic model-building genetic algorithms 330
probabilistic programming languages 281
probabilistic STRIPS operator 260
probimax search 568
progressive unpruning/widening 553
projected Bellman equation 80
projected policy evaluation 80
  projections 224
  projected Bellman equation 225
Q-function 17
Q-learning 31, 227
  difficulty scaling 569
  in Chess 546
QBG 487
QMDP 487
QPOMDP 487
quiescence search 547
R-IAC 130
R-Max 121
randomness 544, 556
RAVE 553
real-time architecture 119
real-time dynamic programming 27
real-time strategy game 558
difficulty scaling 569
knight’s rush 559
seven-roach rush 560
real-world interactions 584
real-world samples 583
realizable 189
regret 180
reinforcement learning 27
in commercial games 570
relation 253, 257
relational
decision list 261
decision tree 260
interpretation 259
learning world models 277
Markov decision process 259
policy 261
policy search 274
POMDP 282
reward function 260
value function 260
relational reinforcement learning 254, 268
relational representation 257
in games 563
representation 9, 589
automatic feature construction 563
combination features 563
expert features 563
hard to learn 564
in Backgammon 544
in Chess 546
relational 563
representation transfer 150
Rescorla-Wagner model 511
reward
local reward 551
reward function 11, 12
reward model
average reward 14
finite horizon 14
infinite horizon 14
rich evaluative feedback 562
risk-sensitive optimization 381
RL-DT 125
robot 579
robust control 381
roll-out 115
rollout 552
rollout analysis 543
rollout policy 553
balanced 554
RSPSA 565
sample complexity of exploration 178
sample efficiency 120
SANE 340
SARSA 31
Sarsa 227
in Neverwinter Nights 571
search 26
self-play 543, 547, 548, 565
semantics of communication 497
semi Markov Decision Problem (SMDP) 297
sequence form 491, 493
sequential decision making 5
sGA 331
shaping 36
simulation 596
simulation balancing 554
SLF-R-Max 124
Snowie 543
sparse sampling 116, 377
speciation 332
SPITI 125
Starcraft 558
state 10
value 15
state abstraction 301
eliminating irrelevant variables 301
funnelling 302
state aggregation 563
state representation
atomic 255
deictic 255
propositional 255
relational 255
state-action value function 17
statistical relational learning 256
stochastic bisimulation homogeneity 302
Stratagus 558
structural credit assignment 8
structure learning 123
structured Bellman backup 262
sub-tree policy 478
substitution 259
subsumption 259
Suicide chess 550
synchronous updating 25
SZ-Tetris 556 558

tabular model 113
targeted exploration 112
task-hierarchy 300
partial-order 316
taxi 123
TD-Gammon 543
vs. humans 543
TD-Leaf 547
exploration 565
TD-learning
in Magic: The Gathering 571
TD-POMDP 495
temporal credit assignment 7
temporal credit assignment problem 27
temporal difference 511
temporal difference learning 29 543
and tree search 547
of value function slope 546
TD(\(\lambda\)) 543
temporal-difference learning 226
GTD2 228
Q-learning 227
Sarsa 227
TD-learning 226
TDC 228
terminal state 13
Tetris 555
NP-hardness 556
TEXPLORER 129
Thompson sampling 377
tile coding 214
time-to-go 478
TOI-Dec-MDP 494
topology 331
training regime 565
database play 547. 566
expert training 548

lesson and practice 566
online game server 548
self-play 543 547. 548 565
tutoring 566
transfer learning 317
transition function 11. 12
Markovian 11
tree search
\(\alpha\beta\)-pruning 551
principal leaf 547
TD-Leaf 547
TreeStrap 548
tutoring 566
TWEANN 331

UCT 117 552 see also Monte-Carlo tree search
exploration 565
narrow path to victory 554
shallow traps 555
underlying MDP 487
underlying POMDP 487
Utile Coordination 462

value approximation 211 223
off-policy 223
online 223
value function
as probability of victory 545
vs. preference function 556. 557
value function approximation
weak performance in Tetris 556
value function decomposition 303
value iteration 23
value of perfect information 128. 362 377
Vexbot 568
Warcraft 558

XCS 337