附录C 中英文对照表

中文	英文	缩写
机器学习基础		
人工智能	Artificial Intelligence	AI
机器学习	Machine Learning	ML
深度学校	Deep Learning	DL
多层感知器	Multilayer Perceptron	MLP
深度神经网络	Deep Neural Networks	DNN
卷积神经网络	Convolutional Neural Network	CNN
循环神经网络	Recurrent Neural Network	RNN
人工神经网络	Artificial Neural Network	ANN
长短期记忆	Long Short-Term Memory	LSTM
单元	Cell	
偏差	Bias	
隐藏状态	Hidden State	
单元状态	Cell State	
隐藏层	Hidden Layer	
批大小	Batch Size	
小批量	Mini-Batch	
整流线性单元	Rectified Linear Unit	ReLU

中文	英文	缩写
指数线性单元	Exponential Linear Unit	ELU
梯度下降	Gradient Descent	
随机梯度下降	Stochastic Gradient Descent	SGD
输出层	Output Layer	
权重	Weight	
引理	Lemma	
步长	Step Size	
步幅	Stride	
超参数	Hyperparameter	
输入	Input	
输出	Output	
初始化	Initialize/Initialization	
更新	Update	
协方差	Covariance	
交叉验证	Cross-Validation	
过度拟合	Overfitting	
欠拟合	Underfitting	
权重衰减	Weight Decay	
集成学习	Ensemble Learning	
自动编码器	Autoencoder	AE
变分自动编码器	Variational Autoencoder	VAE
生成对抗网络	Generative Adversarial Networks	GANs
全连接	Fully-Connected	FC
密集层，亦称全连接层	Dense Layer	
朴素贝叶斯	Naive Bayes	
线性回归	Linear Regression	
折页损失函数	Hinge Loss	
KL 散度	Kullback-Leibler Divergence	KL Divergence
多类别	Multinomial	
独热码	One-Hot	
学习率	Learning Rate	
前向传播	Forward Propagation	
反向传播	Backward Propagation	

中文	英文	缩写
批标准化	Batch Normalization	
分对数	Logit	
对数概率	Log Probability	
线段树	Segment Tree	
张量	Tensor	
早停法	Early Stopping	
数据增强	Data Augmentation	
强化学习基础		
状态	State	
状态集	State Set	
动作	Action	
动作集合	Action Set	
观测	Observation	
轨迹	Trajectory	
智能体	Agent	
奖励	Reward	
环境	Environment	
回报	Return	
转移	Transition	
长期回报	Long-Term Return	
短期回报	Short-Term Return	
探索-利用的权衡	Exploration-Exploitation Trade-Off	
确定性转移过程	Deterministic Transition Process	
随机性转移过程	Stochastic Transition Process	
状态转移矩阵	State Transition Matrix	
基准	Baseline	
部分可观测的	Partially Observable	
完全可观测的	Fully Observable	
立即奖励	Immediate Reward	
累积奖励	Cumulative Reward	
非折扣化的回报	Undiscounted Return	
折扣化回报	Discounted Return	
期望回报	Expected Return	

中文	英文	缩写
起始状态分布	Start-State Distribution	
行动者	Actor	
批判者	Critic	
基于模型的	Model-Based	
无模型的	Model-Free	
基于价值的	Value-Based	
基于策略的	Policy-Based	
既定策略	On-Policy	
新定策略	Off-Policy	
在线策略	On-Policy	
离线策略	Off-Policy	
规划	Planning	
试错过程	Trial-and-Error Process	
自省法	Introspection	
时间差分	Temporal Difference	TD
正向运动学	Forward Kimematics	
反向运动学	Inverse Kinematics	
马尔可夫	Markov	
马尔可夫链	Markov Chain	
马尔可夫性质	Markov Property	
时间同质性	Time-Homogeneous	
时间不同质	Time-Inhomogeneous	
折扣因子	Discount Factor	
赌博机	Bandit	
单臂赌博机	Single-Armed Bandit	
多臂赌博机	Multi-Armed Bandit	MAB
健忘对抗者	Oblivious Adversary	
非健忘对抗者	Non-Oblivious Adversary	
全信息博弈	Full-Information Game	
部分信息博弈	Partial-Information Game	
概率图模型	Probabilistic Graphical Model	
观察变量	Observed Variable	
蒙特卡罗	Monte Carlo	MC

中文	英文	缩写
首次蒙特卡罗	First-Visit Monte Carlo	
每次蒙特卡罗	Every-Visit Monte Carlo	
动态规划	Dynamic Programming	DP
逆矩阵方法	Inverse Matrix Method	
探索和利用	Exploration and Exploitation	
回放缓存	Replay Buffer	
自举	Bootstrap	
穷举法	Exhaustive Method	
非终结	Non-Terminal	
强化学习	Reinforcement Learning	RL
高等强化学习	Advanced Reinforcement Learning	
深度强化学习	Deep Reinforcement Learning	DRL
回合/片段	Episode	
回溯	Backup	
崩溃	Collapse	
截断	Clipped	
贝尔曼方程	Bellman Equation	
贝尔曼期望方程	Bellman Expectation Equation	
贝尔曼最优方程	Bellman Optimality Equation	
贝尔曼最优回溯算子	Bellman Optimality Backup Operator	
批量	Batch	
函数拟合器	Function Approximator	
马尔可夫过程	Markov Process	MP
马尔可夫奖励过程	Markov Reward Process	MRP
奖励函数	Reward Function	
奖励折扣因子	Reward Discount Factor	
马尔可夫决策过程	Markov Decision Process	MDP
有限范围马尔可夫决策过程	Finite-Horizon Markov Decision Process	
部分可观测的马尔可夫决策过程	Partially Observed Markov Decision Process	POMDP
贪心策略	Greedy Policy	
ϵ-贪心	ϵ-Greedy	
后悔值	Regret	

中文	英文	缩写
置信上界	Upper Confidence Bound	UCB
树置信上界	Upper Confidence Bound in Tree	UCT
雅达利游戏	Atari Game	
价值函数	Value Function	
Q 值函数	Q-Value Function	
动作价值函数	Action-Value Function	
在线价值函数	On-Policy Value Function	
最优价值函数	Optimal Value Function	
在线动作价值函数	On-Policy Action-Value Function	
最优动作价值函数	Optimal Action-Value Function	
查找表	Lookup Table	
多项式族	Polynomial Family	
多项式基	Polynomial Basis	
傅立叶基	Fourier Basis	
傅立叶变换	Fourier Transformation	
粗略编码	Coarse Coding	
瓦式编码	Tile Coding	
感知域	Receptive Field	
径向基函数	Radial Basis Function	RBF
决策树	Decision Tree	
最近邻	Nearest Neighbor	
半梯度	Semi-Gradient	
死亡三件套	the Deadly Triad	
过估计	Over-Estimation/Over-Estimate	
欠估计	Under-Estimation/Under-Estimate	
均方误差	Mean Squared Error	MSE
平均绝对误差	Mean Absolute Error	MAE
策略梯度	Policy Gradient	PG
确定性策略	Deterministic Policy	
随机性策略分布	Stochastic Policy Distribution	
确定性策略梯度	Deterministic Policy Gradient	DPG
随机性策略梯度	Stochastic Policy Gradient	SPG
条件概率分布	Conditional Probability Distribution	

中文	英文	缩写
初版策略梯度	Vanilla Policy Gradient	VPG
参数化策略	Parameterized Policy	
伯努利分布	Bernoulli Distribution	
类别分布	Categorical Distribution	
对角高斯分布	Diagonal Gaussian Distribution	
二值化动作策略	Binary-Action Policy	
类别型策略	Categorical Policy	
逐个元素的乘积	Element-Wise Product	
耿贝尔分布	Gumbel Distribution	
耿贝尔-Softmax 函数	Gumbel-Softmax	
耿贝尔-最大化函数	Gumbel-Max	
不可微的	Non-Differentiable	
逆变换	Inverse Transform	
对角高斯策略	Diagonal Gaussian Policy	
累计折扣奖励	Cumulative Discounted Reward	
折扣状态分布	Discounted State Distribution	
转移概率	Transition Distribution	
对数-导数技巧	Log-Derivative Trick	
对数	Logarithm	
将得到的奖励	Reward-to-Go	
偏微分	Partial Derivative	
贯穿时间的反向传播	Backpropagation Through Time	BPTT
莱布尼茨积分法则	Leibniz Integral Rule	
富比尼定理	Fubini's Theorem	
积测度	Product Measure	
可测函数	Measurable Function	
紧致性	Compactness	
被积函数	Integrand	
行为策略	Behaviour Policy	
约等于	Approximately Equivalent	
常规 delta-近似	Regular delta-Approximation	
利普希茨	Lipschitz	
目标网络	Target Network	

中文	英文	缩写
得分函数	Score Function	
路径导数	Pathwise Derivative	
再参数化	Reparametrization	
随机价值梯度	Stochastic Value Gradient	SVG
协方差矩阵自适应	Covariance Matrix Adaptation	CMA
协方差矩阵自适应进化策略	Covariance Matrix Adaptation Evolution Strategy	CMA-ES
爬山法	Hill Climbing	
选择比率	Selection Ratio	
兼容函数近似	Compatible Function Approximation	
优势函数	Advantage Function	
中央处理器	Central Processing Unit	CPU
图形处理器	Graphics Processing Unit	GPU
样本效率	Sample Efficiency	
高样本效率的	Sample-Efficient	
灾难性遗忘	Catastrophic Interference/Forgetting	
元学习	Meta-Learning	
表征学习	Representation Learning	
多智能体强化学习	Multi-Agent Reinforcement Learning	MARL
模拟到现实	Simulation-to-Reality	Sim2Real, Sim-to-Real
信赖域	Trust Region	
共轭梯度	Conjugate Gradient	
自然梯度	Nature Gradient	
变分推断	Variational Inference	VI
专家示范	Expert Demonstrations	
模仿学习	Imitation Learning	IL
交叉熵	Cross Entropy	CE
分层强化学习	Hierarchical Reinforcement Learning	HRL
封建制强化学习	Feudal Reinforcement Learning	
无行动者	Actor-Free	
逆向强化学习	Inverse Reinforcement Learning	IRL
行为克隆	Behavioral Cloning	BC

中文	英文	缩写
学徒学习	Apprenticeship Learning	
从观察量进行模仿学习	Imitation Learning from Observations	IfO/ILFO
高斯混合模型回归	Gaussian Mixture (Model) Regression	GMR
高斯过程回归	Gaussian Process Regression	
因果熵	Causal Entropy	
协变量漂移	Covariate Shift	
复合误差	Compounding Errors	
数据集聚合	Dataset Aggregation	DAgger
无悔的	No-Regret	
动态运动基元	Dynamic Movement Primitives	DMP
单样本的	One-Shot	
最大熵逆向强化学习	Maximum Entropy Inverse Reinforcement Learning	MaxEnt IRL
奖励塑形	Reward Shaping	
生成对抗模仿学习	Generative Adversarial Imitation Learning	GAIL
辨别器	Discriminator	
多模态的	Multi-Modal	
指导性代价学习	Guided Cost Learning	GCL
生成对抗网络指导性代价学习	Generative Adversarial Network Guided Cost Learning	GAN-GCL
极大似然估计	Maximum Likelihood Estimation	MLE
以轨迹为中心的	Trajectory-Centric	
以状态为中心的	State-Centric	
玻尔兹曼分布	Boltzmann Distribution	
配分函数	Partition Function	
重要性采样	Importance Sampling	
对抗性逆向强化学习	Adversarial Inverse Reinforcement Learning	AIRL
互信息	Mutual Information	
时间步	Time Step	
逆向动态模型	Inverse Dynamics Models	
正向动态模型	Forward Dynamics Models	
贝叶斯优化	Bayesian Optimization	BO
从观察量模仿潜在策略	Imitating Latent Policies from Observation	ILPO

中文	英文	缩写
选项框架	Options Framework	
本体感觉	Proprioceptive	
线性二次型调节器	Linear Quadratic Regulator	LQR
极小化极大	Minimax	
从观察量进行行为克隆	Behavioral Cloning from Observation	BCO
正向对抗式模仿学习	Forward Adversarial Imitation Learning	FAIL
动作指导性对抗式模仿学习	Action-Guided Adversarial Imitation Learning	AGAIL
增强逆向动态建模	Reinforced Inverse Dynamics Modeling	RIDM
奖励函数工程	Reward Engineering	
欧氏距离	Euclidean Distance	
时间对比网络	Time-Contrastive Networks	TCN
具象不匹配	Embodiment Mismatch	
概率性运动基元	Probabilistic Movement Primitives	ProMP
核运动基元	Kernelized Movement Primitives	KMP
高斯过程回归	Gaussian Process Regression	GPR
高斯混合模型	Gaussian Mixture Model	GMM
策略替换	Policy Replacement	
残差策略学习	Residual Policy Learning	
基于示范的深度 Q-learning	Deep Q-learning from Demonstrations	DQfD
基于示范的深度确定性策略梯度	Deep Deterministic Policy Gradient from Demonstrations	DDPGfD
标准化 Actor-Critic	Normalized Actor-Critic	NAC
最先进的	State-of-the-Art	SOTA
用示范数据进行奖励塑形	Reward Shaping with Demonstrations	
对比正向动态	Contrastive Forward Dynamics	CFD
内在奖励	Intrinsic Reward	
封建制网络	Feudal Network	FuN
基于族群的训练	Population-Based Training	PBT
通用性	Generality	
多面性	Versatility	
与模型无关的元学习	Model-Agnostic Meta-Learning	
学会学习	Learning to Learn	

中文	英文	缩写
内循环	Inner-Loop	
外循环	Outer-Loop	
元学习者	Meta-Learner	
度量学习	Metric Learning	
元强化学习	Meta-Reinforcement Learning	
小样本学习	Few-Shot Learning	
状态表征学习	State Representation Learning	SRL
描述器	Descriptor	
博弈论	Game Theory	
自我博弈	Self-Play	SP
优先虚拟自我博弈	Prioritized Fictitious Self-Play	PFSP
指导性策略搜搜	Guided Policy Search	GPS
比例-积分-微分	Proportional-Integral-Derivative	PID
现实鸿沟	Reality Gap	
系统识别	System Identification	SI
泛化力模型	Generalized Force Model	GFM
零样本	Zero-Shot	
域自适应	Domain Adaption	DA
渐进网络	Progressive Networks	
动力学随机化	Dynamics Randomization	DR
随机到标准自适应网络	Randomized-to-Canonical Adaptation Networks	RCANs
可扩展性	Scalability	
重要性加权的行动者-学习者结构	Importance Weighted Actor-Learner Architecture	IMPALA
可扩展高效深度强化学习	Scalable, Efficient Deep-RL	SEED
社交树	Social Tree	
多步学习	Multi-Step Learning	
噪声网络	Noisy Nets	
值分布强化学习	Distributional Reinforcement Learning	
分布式贝尔曼算子	Distributional Bellman Operator	
自适应的	Adaptive	
层标准化	Layer Normalization	

中文	英文	缩写
子迭代	Sub-Iteration	
分块对角矩阵	Block Diagonal Matrix	
无穷范式	∞-Norm	
L2 范式	L2-Norm	
模拟	Simulation	
评估/估计	Evaluate	
策略迭代	Policy Iteration	
策略评估	Policy Evaluation	
策略提升	Policy Improvement	
泛化策略迭代	Generalized Policy Iteration	GPI
柔性策略迭代	Soft Policy Iteration	
价值迭代	Value Iteration	
最优性原则	Principle of Optimality	
优先扫描	Prioritized Sweeping	
梯度赌博机	Gradient Bandit	
直接策略搜索	Direct Policy Search	
资格迹	Eligibility Trace	
延迟帧	Lazy-Frame	
选项策略	Policy-over-Action	
选项内置策略	Intra-Option Policy	
时域抽象	Temporal Abstraction	
专注写作	Attentive Writing	
选项内置策略梯度理论	Intra-Option Policy Gradient Theorem	
奖励隐藏	Reward Hiding	
信息隐藏	Information Hiding	
半马尔可夫决策过程	Semi-Markov Decision Process	SMDP
转移策略梯度	Transition Policy Gradients	
重标记	Re-Label	
原始值函数	Proto-Value Functions	PVFs
后见之明目标转移	Hindsight Goal Transitions	
终生学习	Lifelong Learning	
–	Ornstein-Uhlenbeck	OU
斯塔克尔伯格博弈	Stackelberg Game	

中文	英文	缩写
先发优势	First-Mover Advantage	
演算	Roll-Out	
消息传递接口	Message Passing Interfaces	MPI
进程间通信	Inter-Process Communication	IPC
预测者	Predictor	
训练者	Trainer	
	强化学习算法	
探索和利用的指数加权算法	Exponential-Weight Algorithm for Exploration and Exploitation	Exp3
单步 Q-learning	One-Step Q-learning	
多步 Q-learning	Multi-Steps Q-learning	
深度 Q 网络	Deep Q-Networks	DQN
–	Categorical 51	C51
深度确定性策略梯度	Deep Deterministic Policy Gradient	DDPG
优先经验回放	Prioritized Experience Replay	PER
后见之明经验回放	Hindsight Experience Replay	HER
信赖域策略优化	Trust Region Policy Optimization	TRPO
近端策略优化	Proximal Policy Optimization	PPO
分布式近端策略优化	Distributed Proximal Policy Optimizaion	DPPO
–	Actor-Critic	AC
归一化 Actor-Critic	Normalized Actor-Critic	NAC
使用 Kronecker 因子化信赖域的 Actor Critic	Actor Critic Using Kronecker-Factored Trust Region	ACKTR
（同步）优势 Actor-Critic	Synchronous Advantage Actor-Critic	A2C
异步优势 Actor-Critic	Asynchronous Advantage Actor-Critic	A3C
最大化后验策略梯度	Maximum a Posteriori Policy Optimization	MPO
期望最大化算法	Expectation Maximization	EM
拟合 Q 迭代	Fitted Q Iteration	
在线 Q 迭代	Online Q Iteration	
分位数 QT-Opt	Quantile QT-Opt	Q2-Opt
有基准的 REINFORCE	REINFORCE with Baseline	
孪生延迟 DDPG	Twin Delayed DDPG	TD3
柔性 Actor-Critic	Soft Actor-Critic	SAC

中文	英文	缩写
变分信息量最大化探索	Variational Information Maximizing Exploration	VIME
朴素蒙特卡罗搜索	Simple Monte Carlo Search	
蒙特卡罗树搜索	Monte Carlo Tree Search	MCTS
多智能体 Q-learning	Multi-Agent Q-learning	
多智能体深度确定性策略梯度	Multi-Agent Deep Deterministic Policy Gradient	MADDPG
截断 Double-Q Learning	Clipped Double-Q learning	
分布式深度循环回放 DQN	Recurrent Replay Distributed DQN	R2D2
回溯-行动者	Retrace-Actor	
分位数回归 DQN	Quantile Regression DQN	QR-DQN
战略专注作家	Strategic Attentive Writer	STRAW
选项批判者	Option-Critic	
MAXQ 分解	MAXQ Decomposition	
层次抽象机	Hierarchical Abstract Machines	HAMs
使用离线策略修正的分层强化学习	Hierarchical Reinforcement Learning with Off-Policy Correction	HIRO
细粒度动作重复	Fine Grained Action Repetition	FiGAR
通用价值函数逼近器	Universal Value Function Approximators	UVFAs
GPU/CPU 混合式异步优势 Actor-Critic	Hybrid GPU/CPU Asynchronous Advantage Actor-Critic	GA3C
其他		
个人主页	Homepage	
章节	Chapter	
小节	Section	
简介	Introduction	
代码库	Repository	