



价值函数

策略

Actor-Critic

(QAC, A2C,
A3C等)

QT-Opt

基于价值

(Q-learning, DQN等)

基于策略

(REINFORCE, CE方法)