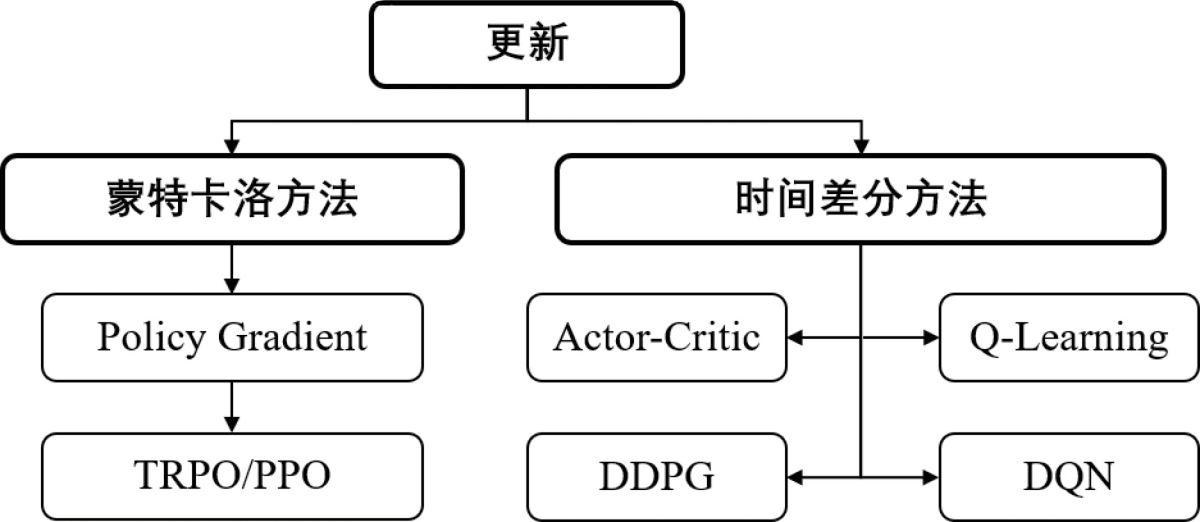
更新
蒙特卡洛方法
时间差分方法
Policy Gradient
Actor-Critic
Q-Learning
TRPO/PPO
DDPG
DQN