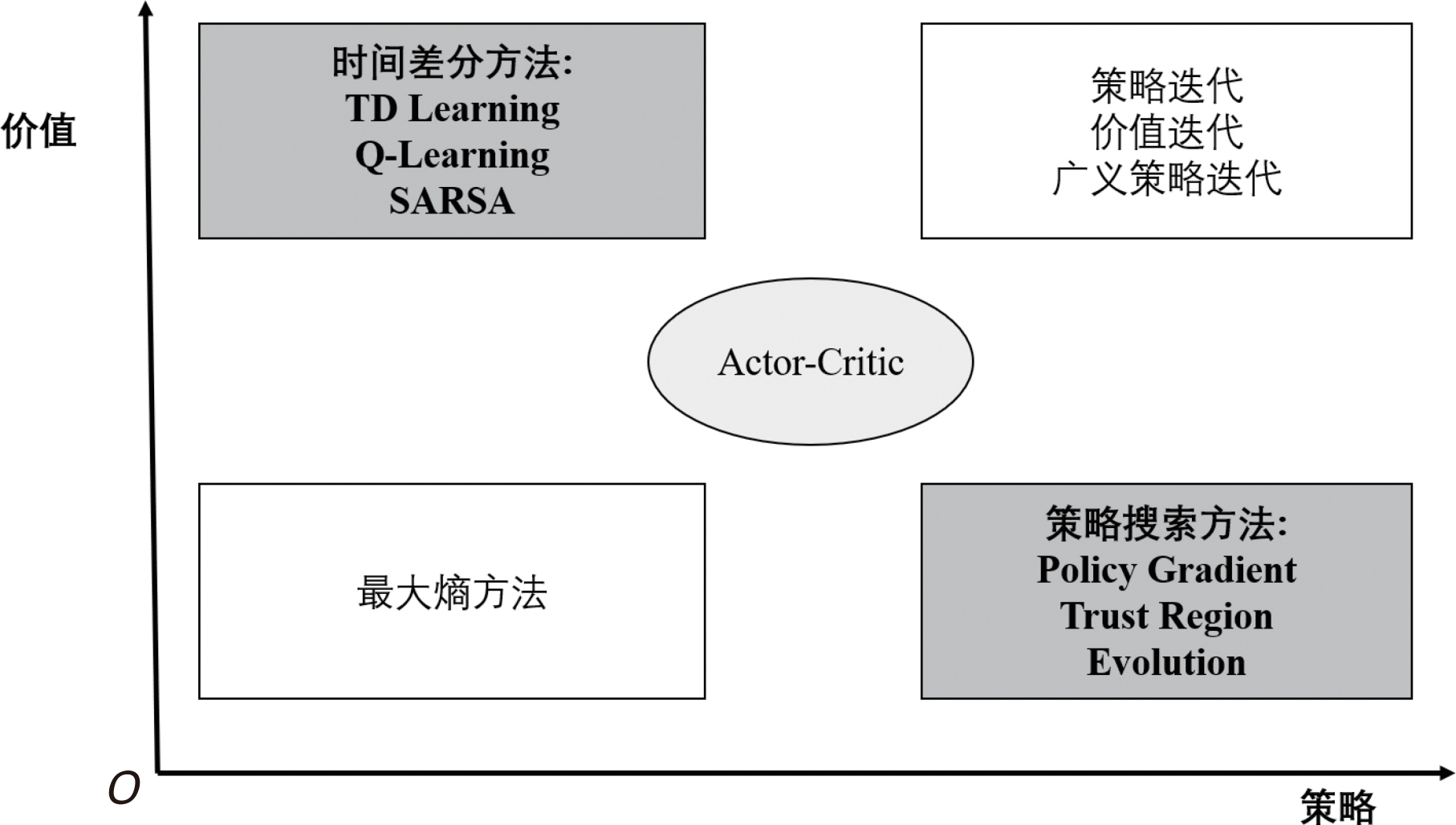
价值
时间差分方法:
TD Learning
Q-Learning
SARSA
策略迭代
价值迭代
广义策略迭代
Actor-Critic
最大熵方法
策略搜索方法:
Policy Gradient
Trust Region
Evolution
O
策略