1) 策略对比:

DQN	Double DQN	Duel DQN	DDPG	A3C
离策略	离策略	离策略	离策略	离策略,异步训练
原始輸入 ▼ Q网络 ▼ 行为Q值	原始输入 ▼ Q网络 ▼ 行为Q值	原始输入 feature网络 V输出 fabrus fab	原始输入 集略网络 集略网络 action	原始输入 策略网络 V值 action
需要	需要	需要	需要	不需要
greedy-e	greedy-e	greedy-e	ou 过程	随机策略采样
最小化 target Q 和 Q 的误差	最小化 target Q 和 Q 的误差	最小化 target Q 和 Q 的误差	最大化 action 和 action 对应的 Q 值 的联合,最小化 Q 值和 target Q 值的 差	最大化 action 和 V 值的联合,最小化 V 值和 target V 值 的差

2) 策略对比:

