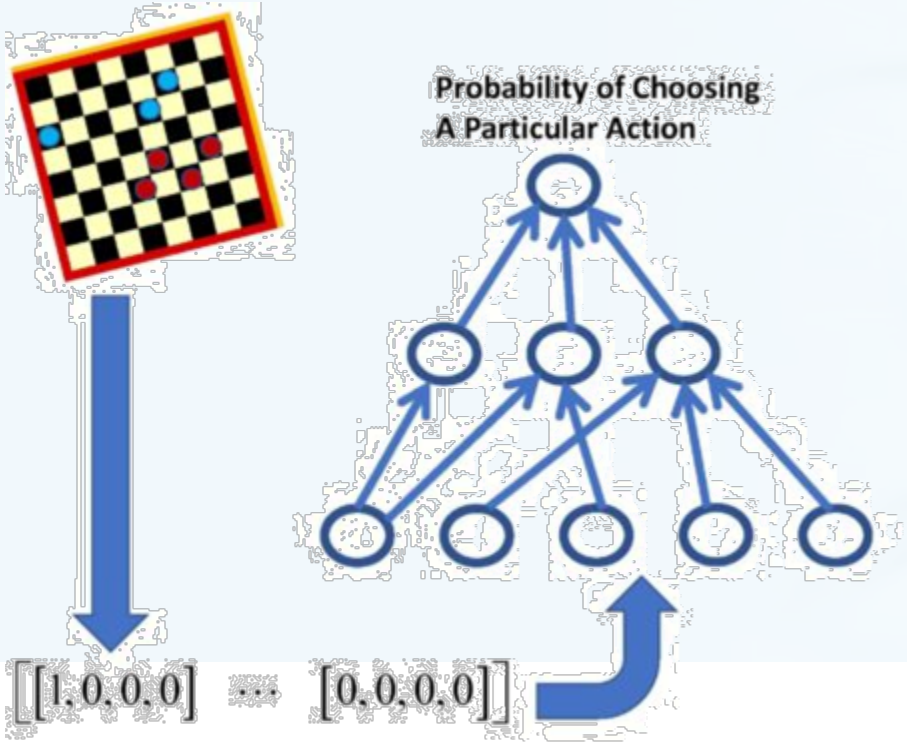
人工智能与信息社会

基于神经网络的智能系统II：价值判断 Q函数

陈斌 北京大学 gischen@pku.edu.cn

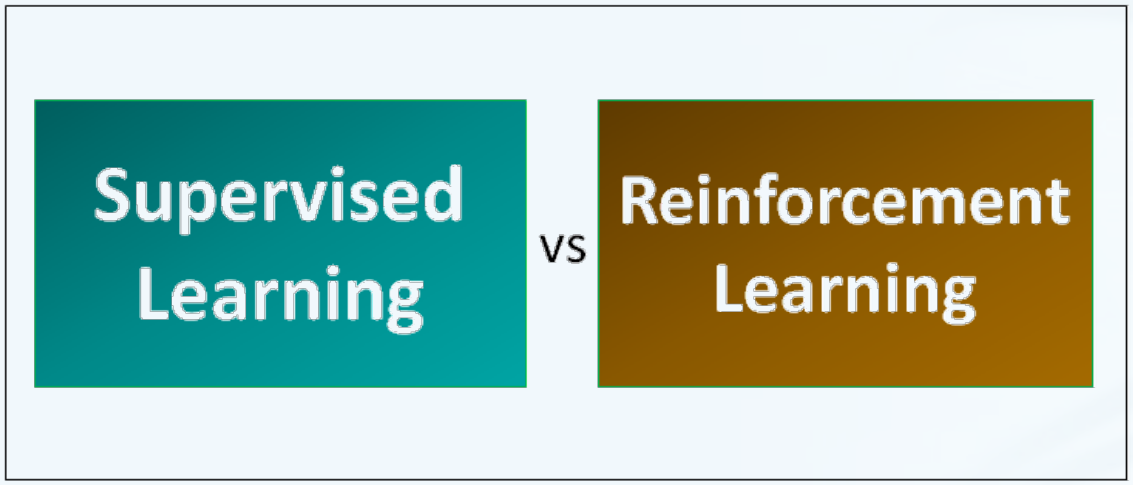
策略

- 从状态集（所有可能出现的状态）到动作集（所有可能采取的动作）的一个对应关系。



目标：求得最佳策略

- 与手写数字识别不同，在强化学习中我们不关心把当前的状态分为什么类型，而是关心它能否执行最佳动作。



监督学习　　强化学习

判断状态

- **状态值函数V**

  只和状态相关，用于对某个局面状态进行估值。

- **状态动作函数Q**

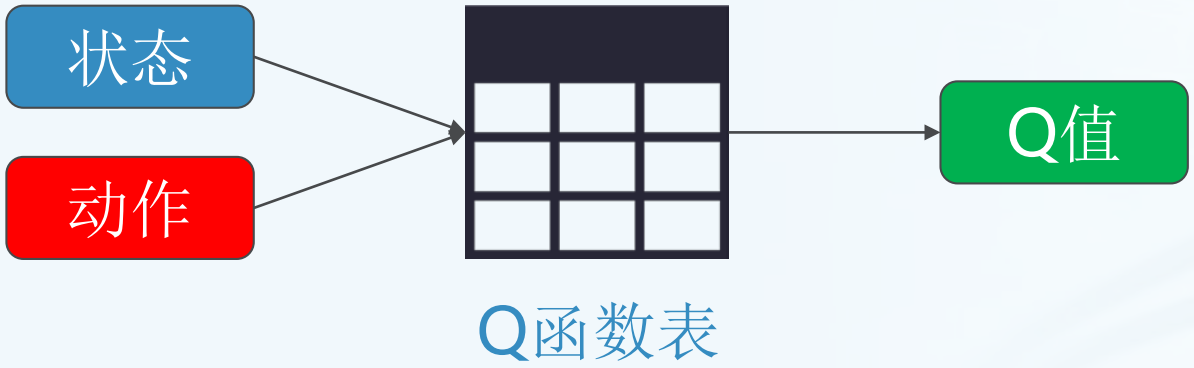
  和状态以及在该状态下采取的动作相关，用于对某个局面状态下采取某个动作进行估值。

Q-Learning

- 强化学习中一种常用算法。
- 基于状态动作函数Q，如果知道了某一状态下每个动作的估值，那么就可以选择估值最好的一个动作去执行了。

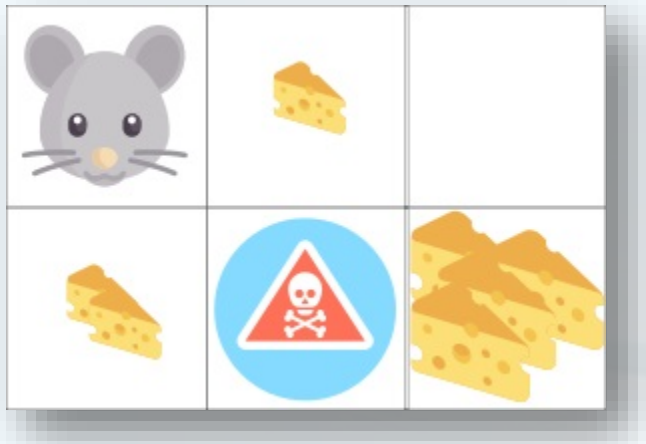
简单的Q函数表（Q-Table）

- Q函数表中行表示状态，列表示动作，表中的值表示特定状态下执行某动作的评估值Q。
- 主体通过不断更新并查找该表，找到当前状态回报最高的动作执行。



简单的Q函数表（Q-Table）

示例

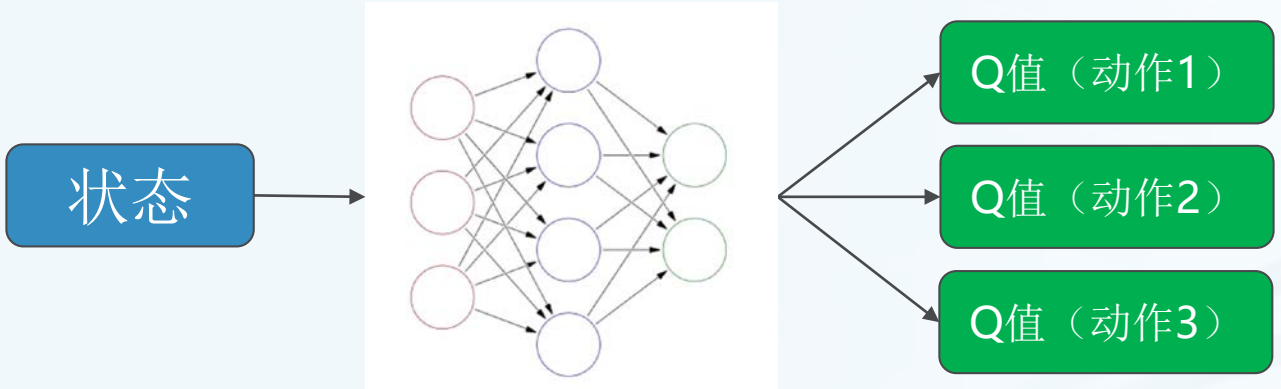


某个策略的Q函数表

状态\动作	上	下	左	右
开始点	0	20	0	10
一小块奶酪	0	-100	1	2
空白	0	100	10	0
两小块奶酪	5	0	0	-100
毒药	0	0	0	0
一堆奶酪（终点）	0	0	0	0

基于神经网络计算Q函数

- 对于复杂的状态，无法用表格表示，可使用神经网络对Q函数进行建模，其输入为状态，输出为各个动作的评估值。还是选取最高的动作执行。



Q神经网络

总结

- Q-Learning算法通过学习获得一个状态动作函数（Q函数）
- 不直接决定主体该采取什么决策，而是提供一个估值参考。
- 如果Q函数较优，可以直接取最大价值来决定动作。