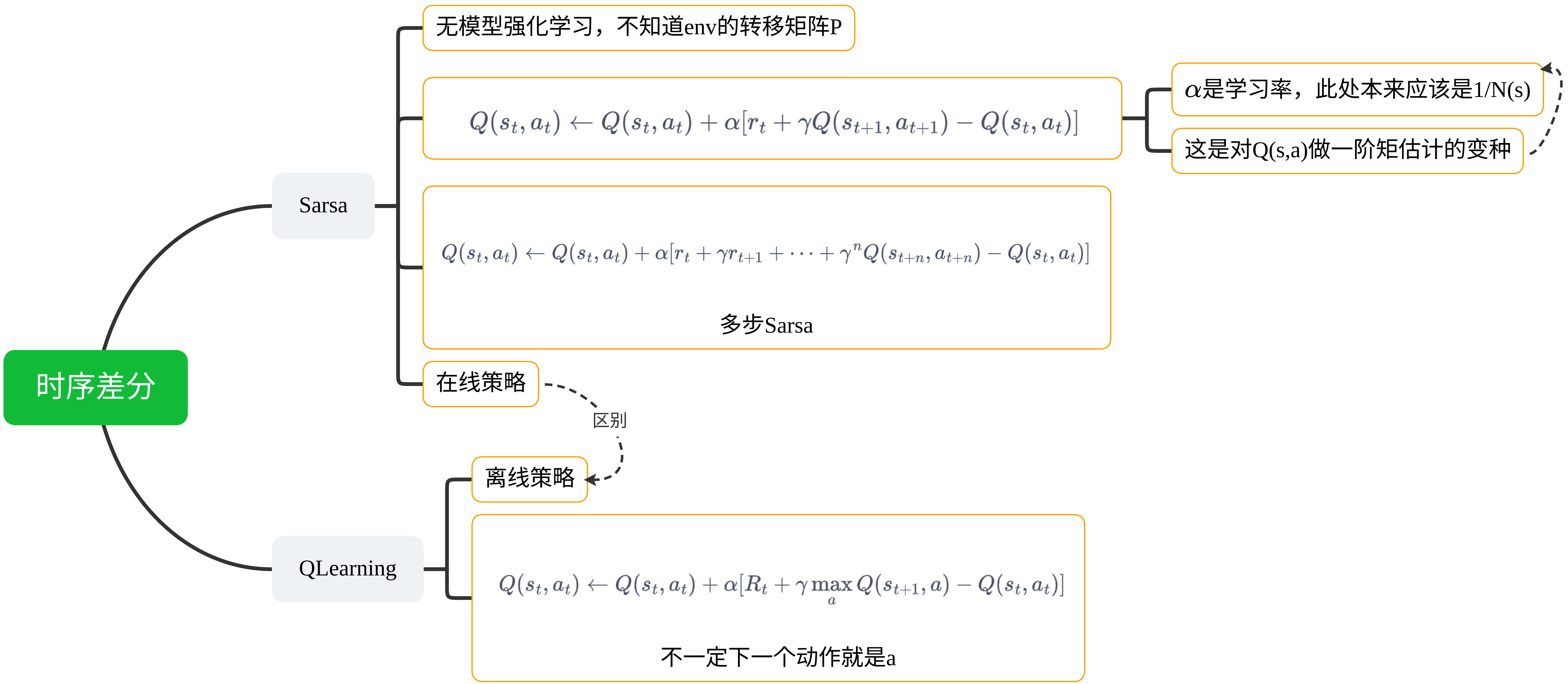

# 时序差分

- Sarsa
  - 无模型强化学习，不知道env的转移矩阵P
  - $Q(s_t,a_t) \leftarrow Q(s_t,a_t)+\alpha[r_t+\gamma Q(s_{t+1},a_{t+1})-Q(s_t,a_t)]$
    - $\alpha$是学习率，此处本来应该是1/N(s)
    - 这是对Q(s,a)做一阶矩估计的变种
  - $Q(s_t,a_t) \leftarrow Q(s_t,a_t)+\alpha[r_t+\gamma r_{t+1}+\cdots+\gamma^n Q(s_{t+n},a_{t+n})-Q(s_t,a_t)]$

    多步Sarsa
  - 在线策略
- QLearning
  - 离线策略
  - $Q(s_t,a_t) \leftarrow Q(s_t,a_t)+\alpha[R_t+\gamma \max_a Q(s_{t+1},a)-Q(s_t,a_t)]$

    不一定下一个动作就是a

区别

作者：@ry664663 | 来自：知犀思维导图