



@author: natasha_yang
@e-mail: ityangna0402@163.com

写代码思路

1. 确定使用何种环境,弄清楚环境的接口,以及使用何种policy来计算价值
 2. 根据第一步的计划,初始化环境和各种参数,比如传统的(Q-table)+DL的搭建
 3. 搭建大框架
- 玩多少次(episode)
- 环境重置
- 刷新环境
- 根据当前状态选择action
- 根据选择的action更新状态并计算当前状态的自身价值
- get下一状态的价值计算出当前状态的价值

