Sarsa算法与Q-learning算法实例-Windy Gridworld

作为机器人移动的一个变种

在如图的10x7的长方形格子世界，标记有一个起始位置S以及一个终止目标位置G

格子下方的数字代表对应列中一定强度的风

当个体进入该列中的某一个格子时，会按照如图所示的箭头所示的方向移动数字所代表的格数

但是格子是有边界的，不会出界

个体：不知道这个世界构造以及风，也不知道边界在哪，不知道格子之间的对应关系，

不记得初始位置与终止位置，但是会记得曾经走过的格子，进入一个格子会记得是那么时候来过

并且动作只有上下左右，每移动一步，如果不是进入目标位置都会给予惩罚

直到进入目标位置获得奖励为0后永久停留在该位置

目标，个体如何能够快速的从启示位置抵达终点位置，求解一个策略