Два бонуса за action =6 or action=7

Правильная стратегия вправо и стрелять!!!

n\_episodes = 100 # количество эпизодов lapan = 20

alpha = 0.5 #learning rate sayon - 0.5 больш - 0.9 Lapan = 0.2

gamma = 0.9 #discount factor sayon - 0.9 больш - 0.5 lapan = 0.9

epsilon = 0.7 #e-greedy sayon - 0.3 больш - 0.7 lapan = = 1.0 (100% random actions)

bonusrewardsize = 10 # for fire - action 6

if action==6 or action==7:

bonusreward[agent\_id] += bonusrewardsize

#динамический epsilon - только при большом количестве эпизодов имеет смысл!!!

if e % 15 == 0:

epsilon += (1 - epsilon) \* 10 / n\_episodes

print("epsilon = ", epsilon)

get\_stats()= {'battles\_won': 30, 'battles\_game': 30, 'battles\_draw': 0, 'win\_rate': 1.0, 'timeouts': 0, 'restarts': 0}

actions\_history= [5, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 4, 1, 5, 1, 6, 1, 6, 1, 6, 1, 6, 1, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7]

Average reward = 20.235250538793096

ПОВТОРНОЕ ОБУЧЕНИЕ:

get\_stats()= {'battles\_won': 30, 'battles\_game': 30, 'battles\_draw': 0, 'win\_rate': 1.0, 'timeouts': 0, 'restarts': 0}

actions\_history= [3, 3, 3, 3, 3, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6]

Average reward = 20.195177801724135