

# 强化学习

状态空间由环境设备状态，和任务队列里面的任务的状态投影。

动作就是就绪任务队列里面的任务的编号。

Environment里要有，任务队列，设备队列。观察函数（生成状态），get\_action函数生成动作的

## State有关:

pending\_tasks: 每个元素都是一个五元组，包括任务节点本身,任务所属设备，任务的到达时间，任务在AppDAG图内的ID和任务所属的AppID

ready\_tasks: 每个元素是一个五元组，包括任务节点本身,任务所属设备，任务的到达时间，任务在AppDAG图内的ID和任务所属的AppID

## TaskDAG有关:

TaskDAG中的每一个node是一个二元组，第一个值是node的index第二个为一个TaskNode节点

Task里面再加一个调度rank队列，每次scheduling时向里面加入任务以记录调度顺序

## EnvInit有关:

Env中的tasks：一个四元组

```
env.tasks.append((task_dag, task_type, 0, arrival_time))
# 第一个元素是任务的DAG图，第二个元素表示任务的所属设备，第三个是属于哪个设备，第四给变量表示为任务的到达时间
```

Env中的Devices:

其中每个元素是一个三元组，第一个元素是Device类，第二个元素是Device的类别，第三个元素表示Device的ID

```
self.devices.append((Device(tuple(args.coordinates1), args.computing_speed1, args.channel_gain1,
                             available_time=args.available_time1), device_type, len(self.devices) + 1))
```

Env中的running\_tasks

self.running\_tasks = [] #每个元素是一个五元组，包括任务节点本身,任务执行的设备，任务的完成执行时间，任务在AppDAG图内的ID和任务所属的AppID

```
self.running_tasks.append((task_node, offload_device, state.cur_time + running_time, node_id, appID))
print("执行的任务")
print(f"task_node: {task_node}\n"
      f"app_device: {app_device}\n"
      f"arriving_time: {arriving_time}\n"
      f"offload_device: {offload_device}\n"
      f"finish_time: {state.cur_time + running_time}\n"
      f"node_id: {node_id}\n")
```

## Device有关

```
self.coordinates = coordinates # 设备当前位置
self.computing_speed = computing_speed # 设备计算速度
self.channel_gain = channel_gain # 设备信道增益
self.available_time = available_time # 可获得时间初始化为 0,表示一开始为空闲的
```

## 奖励设置:

## 存在的问题

1.任务没有运行完成之前, ready\_tasks中的任务可能和running\_tasks中任务重, 已经在running\_tasks的任务不应该再在read\_tasks中出现。目前的解决方案是再加一个task\_node\_running\_flag数组标识。

2.第二个问题就是循环可能会空转也就是被依赖任务没有被完成所以后续任务无法被调度只有等时间慢慢过去等到完成了才有新的任务被调度也就是说在强化学习的过程中很多时候action中没有动作(可以设置一下奖励这个地方)。

## 临时结果:

FIFS: 200

```
第51个任务的到达时刻是165的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-165
第52个任务的到达时刻是184的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-184
第53个任务的到达时刻是184的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-184
第54个任务的到达时刻是199的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-199
第55个任务的到达时刻是199的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-199
完成的任务数22
平均响应时间:100.681818181819
seed:4
```

随机:

```
Run Envinit x
第44个任务的到达时刻是181的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-181
第45个任务的到达时刻是181的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-181
第46个任务的到达时刻是195的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-195
第47个任务的到达时刻是195的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-195
完成的任务数19
平均响应时间:98.52631578947368
seed:4

Process finished with exit code 0
```

强化学习第一个大轮次结果:

```
第57个任务的到达时刻是188的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-188
第58个任务的到达时刻是196的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-196
第59个任务的到达时刻是196的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-196
*****第0次(系统共运行200时间)训练结果*****
完成的任务数26
平均响应时间:80.6923076923077
总奖励为:154520
*****第1次(系统共运行200时间)训练结束*****
```

强化学习第二个大轮次结果:

```
Threads & Variables Console
第57个任务的到达时刻是188的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-188
第58个任务的到达时刻是196的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-196
第59个任务的到达时刻是196的完成时刻是0, 响应时间(任务完成时间-任务到达时间)是-196
*****第1次(系统共运行200时间)训练结果*****
完成的任务数25
平均响应时间:82.0
总奖励为:156710
*****第2次(系统共运行200时间)训练结束*****
Process finished with exit code 0
```

强化学习最好的结果:

```
*****第149次(系统共运行200时间)训练结果*****
完成的任务数:27
平均响应时间:72.37037037037037
总奖励为:204160
*****第150次(系统共运行200时间)训练结果*****
```