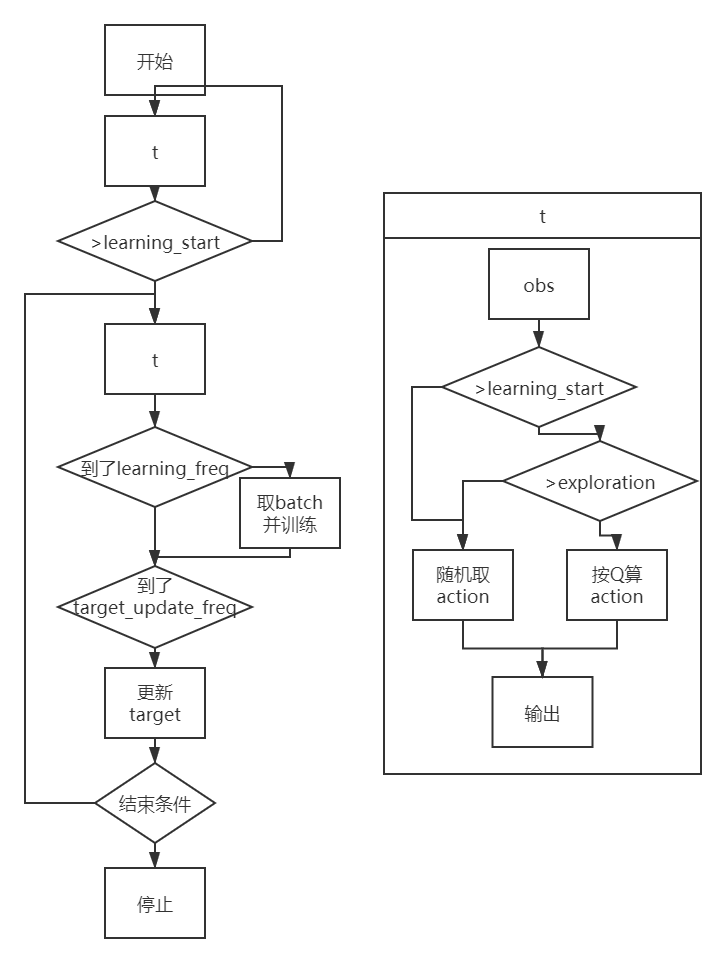
经过尝试发现bus的电压很难超过1.25的临界值……太难了。

找了网上一个参考：https://github.com/transedward/pytorch-dqn。准备以这个为基础改出来我们自己的DQN。



不过这个程序还是有几个东西需要注意：

1. Env我们没有现成的，需要自己打包一个env出来，主要是和pandapower交互，两个函数， reset和step。另外，几个attributes，比如num\_action这种。
2. replay\_buffer这个是用来存训练模型的样本信息的buffer，可能需要改改，但是不需要大改
3. DQN原始输入是图片，改成输入我们的状态空间（刚开始用全量的吧）
4. 电网仿真的话，有更多的初始情况，并且结束的更快。需要考虑到。
5. 需要考虑stopping\_criterion怎么给
6. 思言说最后要搞一个pickle存训练的信息，可能最后弄吧

另外还有一些思考：

这里用到了一个exploration的参数，用来决定是随机取action还是根据Q算action,随时间变化。

用的也是target的网络，我觉得target应该是DQN必备的。

wrapper没搞懂到底是干啥的，可能是用来存log的

reply\_buffer负责搞随机采样，采样是有放回的随机采，我感觉可能会带来样本使用率不均衡的问题，但是思言说没关系。

最后：虽然程序的大致逻辑我看懂了，但是还是有很多细节我没有仔细探究。明天看。

工作流程：

|  |  |  |
| --- | --- | --- |
| 工作点 | 备注 | 工作量 |
| 配置环境并试用Gym的强化学习代码 | anaconda整体迁移内网。**不好评估会不会有问题。** | 中 |
| Env | reward, step和reset三个函数。一些其他属性。 | 高 |
| 状态空间与动作空间 | **怎样提取？是否需要one-hot化？状态空间能否压缩？** | 中 |
| Replay Buffer | 就是loader。用来储存、更新和调用样本。 | 中 |
| schedule | 用来调整DQN的exploitation和exploration的比例。 | 极低 |
| DQN model | 实际的NN模型，从输入图片改成输入action。 | 低 |
| DQN learn | 主程序，需要整体迁移 | 高 |
| 主程序与搜索超参 | 完成主程序。  **是否需要搜索超参？** | 中 |
| 阅读结果并解释 | 从电力系统角度分析实验结果。 | 中 |
| 与美国院讨论 |  | 低 |