amp_network_builder

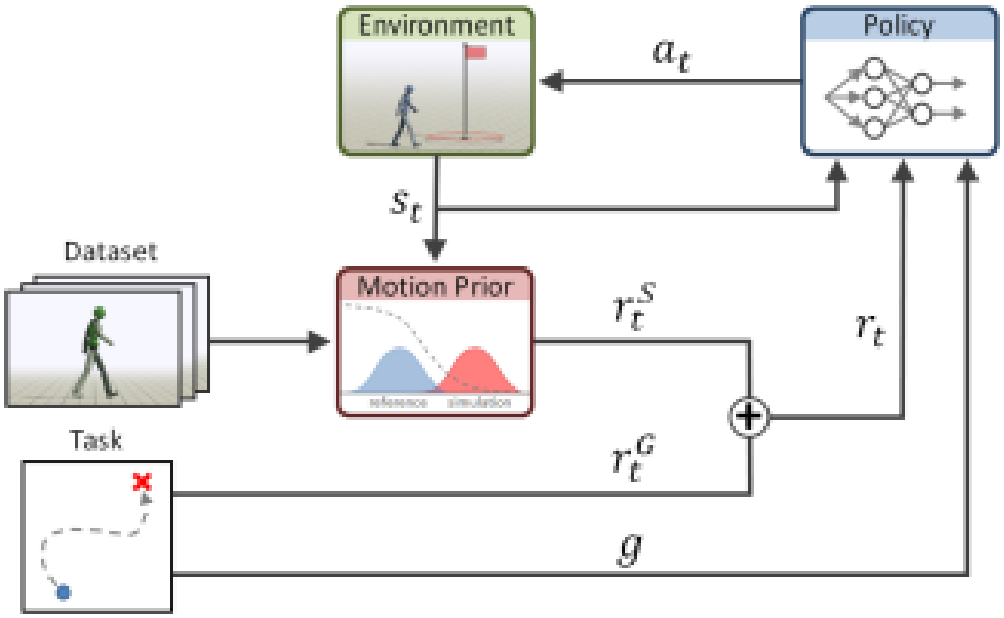
network builder 根据训练配置文件中的网络结构搭建神经网络

amp_network_builder

1. 继承自network_builder.A2CBuilder, network_builder是rl_games里面的内置class. A2C全称Advantage Actor Critic, 是一种actor critic方法，在计算gradient的时候减去一个baseline
2. *network* 是里面的一个嵌套类，同样，继承自network_builder.A2CBuilder.Network
   如果网络用于连续动作空间的任务且不需要学习动作标准差（learn_sigma为False），则初始化一个参数 sigma，用于控制动作的标准差。
   调用 _build_disc 函数，用于构建一个用于区分环境状态的神经网络。

   这个network的意义看起来就是为了搭建GAIL的discriminator.
   a. load() load discriminator 里面的参数 units, activation, initializer
   b. eval_critic(obs) **critic**
      这个是用来给**CRITIC**打分的
   c. eval_disc(amp_obs) **actor**
      评估演员部分的性能，返回判别器网络的输出
   d. _build_disc(input_shape) 搭建了一个discriminator的二分类器，用来判断这个是由生成器生成的还是从motion data里面取出来的。这个网络是由MLP搭的，最后加一个Linear层，使得最后只有一个node

3. build
   最后，AMPBuilder 类定义了一个 build 方法，用于构建 Network 类的实例，该实例代表整个神经网络。这个方法返回一个 Network 对象。



5.1 Imitation from Observations

The original formulation of GAIL requires access to the demonstrator's actions [Ho and Ermon 2016]. However, when the demonstrations are provided in the form of motion clips, the actions taken by the demonstrator are unknown, and only states are observed in the data. To extend GAIL to settings with state-only demonstrations, the discriminator can be trained on state transitions $D(\mathrm{s}, \mathrm{s}')$ instead of state-action pairs $D(\mathrm{s}, \mathrm{a})$ [Torabi et al. 2018],

$$\arg\min_D -\mathbb{E}_{d^{\mathcal{M}}(\mathrm{s},\mathrm{s}')}\left[\log\left(D(\mathrm{s},\mathrm{s}')\right)\right] - \mathbb{E}_{d^{\pi}(\mathrm{s},\mathrm{s}')}\left[\log\left(1 - D(\mathrm{s},\mathrm{s}')\right)\right]. \tag{5}$$

$d^{\mathcal{M}}(\mathrm{s}, \mathrm{s}')$ and $d^{\pi}(\mathrm{s}, \mathrm{s}')$ denote the likelihood of observing a state transition from s to s' in the dataset $\mathcal{M}$ and by following policy π respectively. Note that if the demonstrator is different from the agent (e.g. a human actor), the observed state transitions may not be physically consistent for the agent, and therefore impossible for the agent to perfectly reproduce. Despite this discrepancy, we show that the discriminator still provides an effective objective for imitating a wide range of behaviors.