

Reinforcement Learning

强化学习基础

中山大学计算机学院

目录

1. 深度 Q 学习 (Deep Q-Learning, DQN)
2. 实验任务

3.2 DQN

- Replay Buffer

full Q-learning with replay buffer:

+ samples are no longer correlated

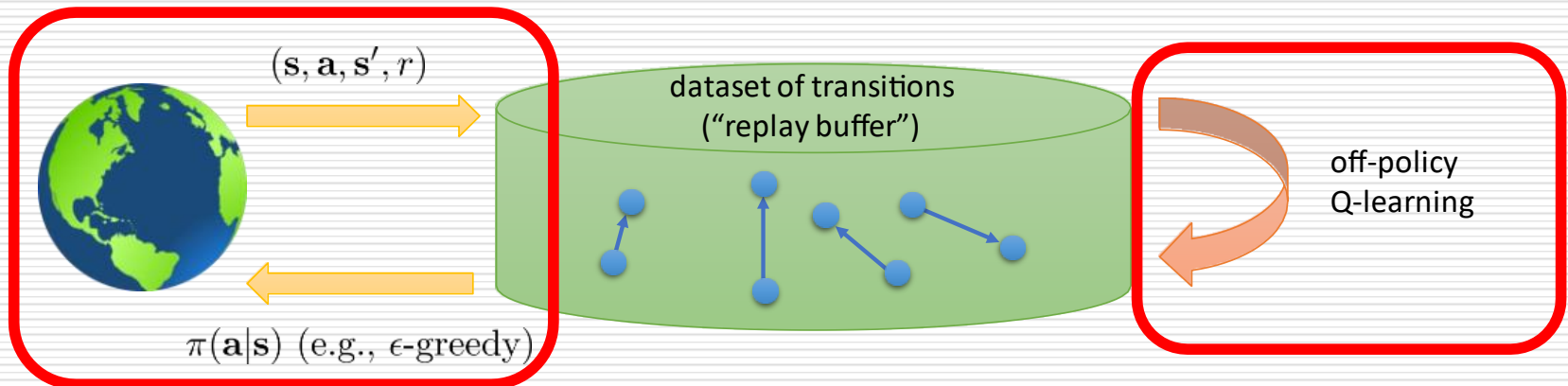
1. collect dataset $\{(s_i, a_i, s'_i, r_i)\}$ using some policy, add it to \mathcal{B}
2. sample a batch (s_i, a_i, s'_i, r_i) from \mathcal{B}
3. $\phi \leftarrow \phi - \alpha \sum_i \frac{dQ_\phi}{d\phi}(s_i, a_i)(Q_\phi(s_i, a_i) - [r(s_i, a_i) + \gamma \max_{a'} Q_\phi(s'_i, a')])$

+ multiple samples in the batch (low-variance gradient)

but where does the data come from?

need to periodically feed the replay buffer...

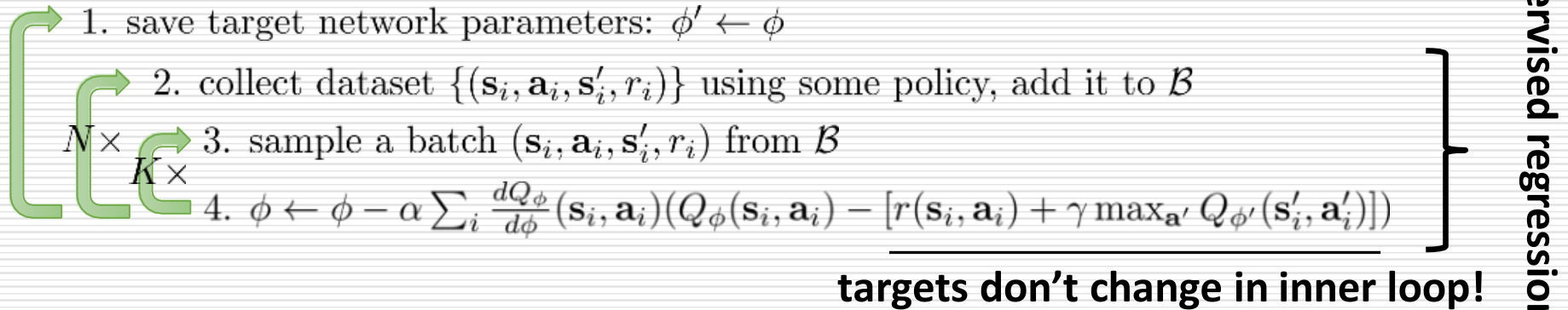
K = 1 is common, though larger K more efficient



3.2 DQN

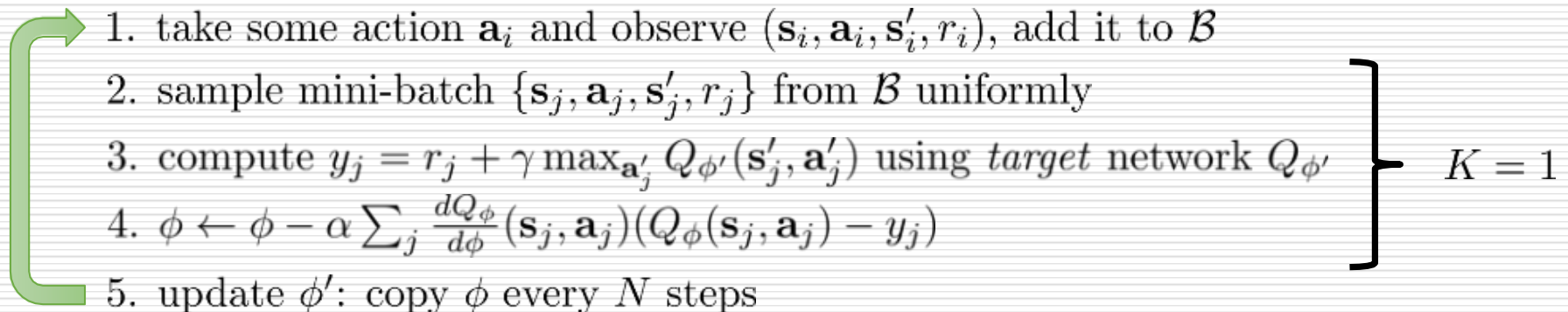
- **Target Networks**

Q-learning with replay buffer and target network:

- 
1. save target network parameters: $\phi' \leftarrow \phi$
 2. collect dataset $\{(\mathbf{s}_i, \mathbf{a}_i, \mathbf{s}'_i, r_i)\}$ using some policy, add it to \mathcal{B}
 3. sample a batch $(\mathbf{s}_i, \mathbf{a}_i, \mathbf{s}'_i, r_i)$ from \mathcal{B}
 4. $\phi \leftarrow \phi - \alpha \sum_i \frac{dQ_\phi}{d\phi}(\mathbf{s}_i, \mathbf{a}_i) (Q_\phi(\mathbf{s}_i, \mathbf{a}_i) - \underbrace{[r(\mathbf{s}_i, \mathbf{a}_i) + \gamma \max_{\mathbf{a}'} Q_{\phi'}(\mathbf{s}'_i, \mathbf{a}'_i)]}_{\text{targets don't change in inner loop!}})$

- **Deep Q-Network(DQN)**

“classic” deep Q-learning algorithm:

- 
1. take some action \mathbf{a}_i and observe $(\mathbf{s}_i, \mathbf{a}_i, \mathbf{s}'_i, r_i)$, add it to \mathcal{B}
 2. sample mini-batch $\{\mathbf{s}_j, \mathbf{a}_j, \mathbf{s}'_j, r_j\}$ from \mathcal{B} uniformly
 3. compute $y_j = r_j + \gamma \max_{\mathbf{a}'} Q_{\phi'}(\mathbf{s}'_j, \mathbf{a}'_j)$ using *target* network $Q_{\phi'}$
 4. $\phi \leftarrow \phi - \alpha \sum_j \frac{dQ_\phi}{d\phi}(\mathbf{s}_j, \mathbf{a}_j) (Q_\phi(\mathbf{s}_j, \mathbf{a}_j) - y_j)$
 5. update ϕ' : copy ϕ every N steps

3 实验任务及报告提交要求

实验任务

- 在gym库的cartpole环境中实现DQN算法。

报告提交要求

- 提交一个压缩包。压缩包命名为：“学号_姓名_作业编号”，例如：
20220525_张三_实验11。
- 压缩包包含三部分：code文件夹和实验报告pdf文件
 - Code文件夹：存放实验代码
 - Pdf文件格式参考发的模板
- 如果需要提交新版本，则在压缩包后面加_v1等。如“学号_姓名_作业编号_v1.zip”，以此类推。
- 截止日期：2022年7月13日24点
- 提交邮箱：zhangyc8@mail2.sysu.edu.cn