💡

InferNet For Delayed Reinforcement Tasks: Addressing the Temporal Credit Assignment Problem

论文试图解决什么问题？

CAP：

Solving the temporal CAP is especially important for delayed reinforcement tasks [2], in which a reward rt obtained at time t, can be affected by all previous actions, a0, a1, ..., at−1, at and thus we need to assign credit or blame to each of those actions individually。

这是否是一个新的问题？

todo

这篇文章要验证一个什么科学假设？

None

有哪些相关研究？如何归类？谁是这一课题在领域内值得关注的研究员？

todo

论文中提到的解决方案之关键是什么？

使用一个神经网络预测中间奖励，从而进行奖励分配。

一系要点：

1. 网络输入：时间步的状态与相应动作
2. 输出：奖励值
3. 对网络作了约束：延迟奖励等于各步的奖励之和：

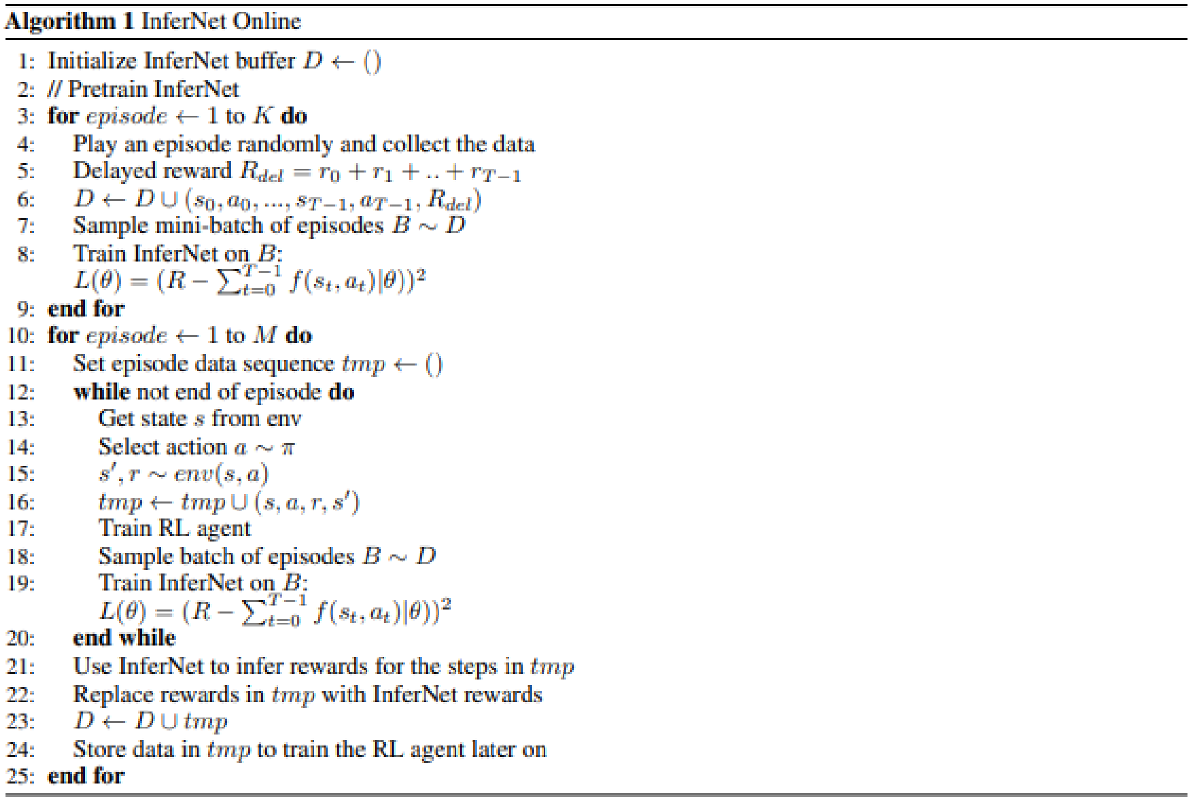
$$R_{del} = f(s_0, a_0|\theta) + f(s_1, a_1|\theta) + ... + f(s_{T-1}, a_{T-1}|\theta)$$

所以Loss为

$$Loss(\theta) = (R_{del} - \sum_{t=1}^{T} f(s_t, a_t|\theta))^2$$

并最小化这个Loss

4. 算法流程：



Algorithm 1 InferNet Online

```
Initialize InferNet buffer $D \leftarrow ()$
// Pretrain InferNet
for $episode \leftarrow 1$ to $K$ do
    Play an episode randomly and collect the data
    Delayed reward $R_{del} = r_0 + r_1 + .. + r_{T-1}$
    $D \leftarrow D \cup (s_0, a_0, ..., s_{T-1}, a_{T-1}, R_{del})$
    Sample mini-batch of episodes $B \sim D$
    Train InferNet on $B$:
    $L(\theta) = (R - \sum_{t=0}^{T-1} f(s_t, a_t)|\theta))^2$
end for
for $episode \leftarrow 1$ to $M$ do
    Set episode data sequence $tmp \leftarrow ()$
    while not end of episode do
        Get state $s$ from env
        Select action $a \sim \pi$
        $s', r \sim env(s, a)$
        $tmp \leftarrow tmp \cup (s, a, r, s')$
        Train RL agent
        Sample batch of episodes $B \sim D$
        Train InferNet on $B$:
        $L(\theta) = (R - \sum_{t=0}^{T-1} f(s_t, a_t)|\theta))^2$
    end while
    Use InferNet to infer rewards for the steps in $tmp$
    Replace rewards in $tmp$ with InferNet rewards
    $D \leftarrow D \cup tmp$
    Store data in $tmp$ to train the RL agent later on
end for
```

论文中的实验是如何设计的？

作者衡量了这On-policy和Off-policy强化学习任务中算法的表现。

用于定量评估的数据集是什么？代码有没有开源？

None

论文中的实验及结果有没有很好地支持需要验证的科学假设？

todo

这篇论文到底有什么贡献？

todo

下一步呢？有什么工作可以继续深入？

todo