

Does Neuron Coverage Matter for Deep Reinforcement Learning? A Preliminary Study

Chen Gong

28 June 2021

此篇文章为：“Does Neuron Coverage Matter for Deep Reinforcement Learning? A Preliminary Study”，于 2020 年发表于 ICSE 的 workshop 上。主要描述了用神经元覆盖实验来进行 RL 算法的测试，看看神经元的覆盖关系和获得较高奖励之间有没有什么联系。文章整体写得比较粗糙，毕竟只是 workshop。

1 Key idea

首先作者认为黑盒测试并不适合于 DeepRL，强化学习测试和普通的机器学习测试相比，难点在于 DeepRL 可以有一系列正确的行为，不同的最有动作（例如，动作是一个合适的范围，而不是某一个准确的值。）当测试 DNN 时，有一个“ground truth”作为评估准确性的 baseline（例如，分类准确类，假阳性）。判断智能体输出的动作是不是最优动作，判断智能体当前的价值函数和最优价值函数之间的差别是非常困难的。

所以，作者考虑白盒测试更适合于 DeepRL。在 DNN 中对于神经元覆盖性测试已经有较好的工作了，而这里可以直接的用于对 DRL 的测试，看看神经元覆盖性测试是否对强化学习有用。本文尝试使用一些常见的测试标准，MC/DC, SS/VB/SV/ VV coverage, neuron boundary coverage, multi-section neuron coverage, strong neuron coverage, 和 top neuron coverage。

本文中将用白盒测试来研究深度网络的覆盖性，在 MountainCar 中，测试了两种不同结构的 DQN 的神经元覆盖性和其获得的奖励之间的关系。**实验结果表明较好的神经元覆盖性，并不意味着在 RL 任务中获得了成功。**此外，更多的研究应该关注于分析覆盖率对 RL 算法有效性的影响，并且通过结合 DeepRL 算法的输出和内部的行为来定义衡量算法有效性的度量标准。

2 Empirical study design

本文感兴趣的是分析 DeepRL 算法整体和神经网络层的神经元覆盖的变换过程，以及它们与累积奖励的相关性（通常使用累积奖励啊来衡量算法的性能）。主要探究了两个问题，

1. RQ_1 : 了解在 DeepRL 系统中，神经元的覆盖变换过程是否因每一层而不同。（这里这个 evaluation 的意思，我实在是没 get 到，希望我理解完全文，能有更好的理解吧。）

2. RQ₂: 在 deep prl 系统中, 神经元覆盖范围和累积奖励之间是否存在相关性? 这个问题的目的是研究更高的累积奖励和更高的覆盖率之间是否有关系。此外, 它还可以指示在训练过程中如何选择参数, 从而通过控制覆盖率, 从而获得更好的回报。

作者选取了两种不同的对 DQN 的实现方式 (不同的网络结构)。模型通过一系列的 iterations, episode 和 stage 来训练和测试。其中, iteration i 表示环境中执行的一个步骤或动作; 一个 episode e 包含最少 200 个 iteration。一个训练 stage 是指一个 training-testing fold, 包含 1000 个 training episode, 每一个 training episode 结束有 20 个 testing episode。所以, 一个 training stage 里面含有 20×1000 个 testing episode。

2.1 分析方法

在 DeepRL 系统的训练和测试过程中, 我们记录了每个 iteration 和 episode 中被激活的神经元。对于系统 s , 每一层为 $l \in s$, 每一个神经元记为 $n \in N_l^s$, 然后第 i 次 iteration 中的, 第 e 个 episode, 的第 n 个神经元的激活情况被记为 $\lambda(n, i, e)$ 。那么, 将用 $|s|$ 来表示系统 s 中神经元的个数, $|l|$ 表示第 l 层中神经元的个数。下面给出几种衡量指标,

1. Neuron Coverage (NC): 第 i 次 iteration 中的, 第 e 个 episode 中, 所有被激活的神经元占有所有神经元的比例,

$$NC(s, i, e) = \frac{\sum_{l \in L} \sum_{n \in N_l^s} \lambda(n, i, e)}{|s|} \quad (1)$$

2. Neuron Layered Coverage (NLC): 考虑的第 l 层神经元被激活的比例,

$$NLC(s, i, e) = \frac{\sum_{n \in N_l^s} \lambda(n, i, e)}{|l|} \quad (2)$$

3. Cumulative Neuron Coverage (CNC): 表示到 iteration i^* 和 episode e^* 为止, 被激活的神经元数量。这里用 $\hat{N}(s, e^*, i^*)$ 来表示, 直到 iteration i^* 和 episode e^* 还没有被激活的神经元, 那么 CNC 被定义为:

$$CNC(s, e^*, i^*) = \frac{\sum_{e \leq e^*} \sum_{i \leq i^*} \sum_{n \notin \hat{N}(s, e^*, i^*)} \lambda(n, i, e)}{|s|} \quad (3)$$

4. Cumulative Neuron Layered Coverage (CNLC): 也是类似的定义,

$$CNLC(s, e^*, i^*, l^*) = \frac{\sum_{e \leq e^*} \sum_{i \leq i^*} \sum_{n \notin \hat{N}(s, e^*, i^*, l^*)} \lambda(n, i, e)}{|l^*|} \quad (4)$$

关于 RQ1 的回答, 用 CNC 和 CNLC 来表示神经元的激活状况, 会如何随着时间而变化。关于 RQ2 的研究, 则是计算 NC 和 NLC 之间的 Pearson 和 Spearman 关联系数, 来区分 coverage 和累积奖励之间是否有关系。

3 实验结果

下图中展示了 CNC 和 CNLC 的实验结果。

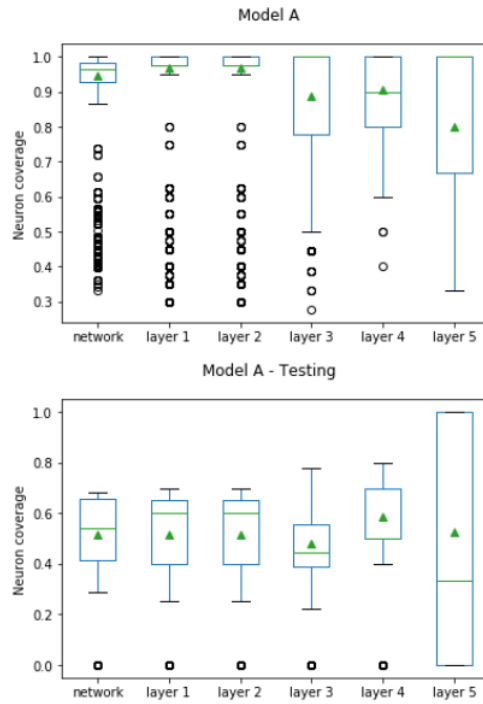


Figure 2: Distribution of cumulative coverage for Model A during the training and testing phases

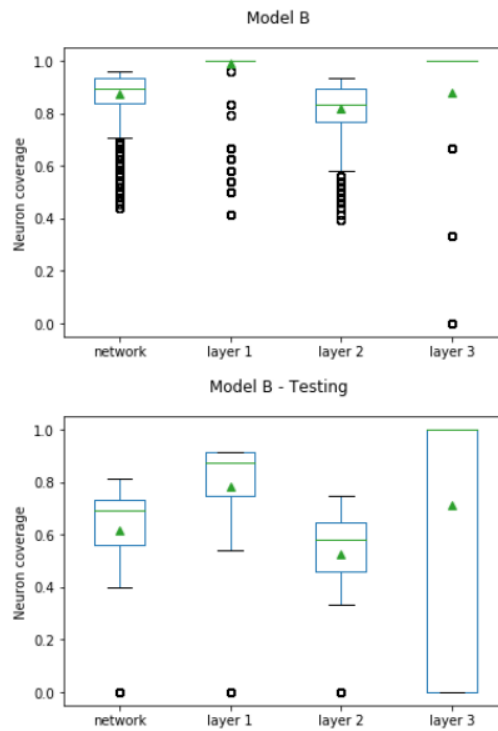


Figure 3: Distribution of cumulative coverage for Model B during the training and testing phases

Training phase

模型 A 和模型 B 中实现了平均为 95% 和 86% 的神经网络覆盖率。在模型 A 中，所有的神经元层都能够在训练中执行所有的神经元。模型 A 中的第 5 层虽然只有 3 个神经元，但两种模型的平均累积覆盖率都较低 (80%)，说明神经元越少并不一定累积覆盖率越高。

详细的研究 CNLC，模型 A 中过程不太稳定，也就是说，就激活一层中所有神经元所需的迭代次数和值的范围而言，这种行为不太一致。模型 B 中所有来自输入层和输出层的神经元都比模型 A 中更快地被激活；然而，这与模型 B 中的隐藏层恰恰相反，其中的进程 (通常) 比模型 A 要慢。那么看到这我理解的前面提出的进化过程，应该指的是，神经元的覆盖性，在训练阶段是如何变化的。

Testing phase 在这两个模型中，我们都可以看到在每一层和网络整体的覆盖率的减少。这里的发现是，模型 A 尽管在训练期间有更好的累积覆盖率，但在测试期间成功率较低且累积覆盖率也较低。也就是说，更好的训练覆盖率并不一定意味着更好的测试覆盖率。此外，在训练期间累积覆盖率的演化模式在测试期间不一定相同。

3.1 Coverage-Reward Correlation

我们可以观察到，在两种模型的不同运行中，最后一层网络的神经元覆盖行为是不稳定的。并且作者观察到，在接近最后一个 episode 的情况下，最有一层的覆盖率接近网络整体的覆盖率。至于神经网络覆盖率和 reward 直接的关系话，得出的结论是没什么关系，也就是并不充分。

4 结论

与 DNN 中的神经元覆盖相比，本文的结果表明，神经元覆盖不足以得出关于 DeepRL 网络设计或结构的实质性结论。最好的覆盖性测试结果是通过广泛的探索来实现的，然而，探索过多并不利于寻找最优策略。因此，对于 DeepRL 来说，神经元覆盖似乎是评估网络早期层的状态空间探索，作为探索不同行为的一种手段。

未来的工作，对神经元覆盖性的评估可以作为有效的度量工具并扩展到其他 DeepRL 场景，来探索一种能够关联覆盖度和最大回报的度量方式 (个人觉得不太可能，应该是作者的一厢情愿)。此外，类似的方法可以用来评估其他基于 dnn 的学习模型。