

Proximal Policy Optimization with Dynamic Clipping

Student: Rishikesh Vaishnav
University of California, San Diego

Mentor: Sicun Gao, PhD
University of California, San Diego

Abstract

Proximal Policy Optimization (PPO), a policy gradient algorithm drawing closely from the theory supporting Trust Region Policy Optimization (TRPO), has emerged as one of the most effective tools in reinforcement learning problems. PPO makes use of a loss function with a clipped importance sampling ratio using a single parameter ϵ . Although PPO shows promising empirical performance, it is vulnerable to problem-specific imbalances in its handling of positive and negative advantages. We investigate one such imbalance, addressing a discrepancy in expected penalty contributions of positive and negative estimators. By precisely calculating this discrepancy and minimizing it before each iteration, we empirically demonstrate that eliminating this discrepancy can improve the overall performance of PPO.

1 Introduction

More fascinating text. Features¹ galore, plethora of promises.

2 This is Another Section

Some embedded literal typset code might look like the following :

```
#include <iostream>
using namespace std;
main()
{
    cout << "Hello world \n";
    return 0;
}
```

Now we're going to cite somebody. Watch for the cite tag. Here it comes [1].

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad

minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Notes

¹Remember to use endnotes, not footnotes!

References

- [1] <https://arxiv.org/abs/1707.06347>
- [2] <https://arxiv.org/abs/1506.02438>
- [3] <https://arxiv.org/abs/1502.05477>