ON THE ESTIMATION BIAS IN DOUBLE Q-LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Double Q-learning is a classical method for reducing overestimation bias, which is caused by taking maximum estimated values in the Bellman operator. Its variants in deep Q-learning paradigm have shown great promise in producing reliable value prediction and further improving the learning performance. However, as shown by prior work, double Q-learning is not fully unbiased and still suffers from a modest underestimation bias. In this paper, we show that such underestimation bias may lead to multiple non-optimal fixed points under an approximated Bellman operation. To address the concerns of converging to non-optimal stationary solutions, we propose a simple and effective approach as a partial fix of underestimation bias, which leverages real returns to bound the target value. We extensively evaluate the proposed method in the Atari benchmark tasks and demonstrate significant improvement over baseline algorithms.

REFERENCES

A APPENDIX

You may include other additional sections here.