

On the Variance of the Adaptive Learning Rate and Beyond

備考

著者

Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, Jiawei Han

掲載

arXiv:1908.03265 [cs, stat]

Abstract

学習率ウォームアップは、RMSpropやAdamのような適応的確率最適化アルゴリズムにおいて、学習の安定化、収束の加速、汎化の改善において顕著な成功を収めている。ウォームアップの背後にある理論を追求し、適応的な学習率の問題点、すなわち、その分散が初期段階で問題なく大きいことを明らかにし、ウォームアップが分散削減技術として働くと推定する。この仮説を検証するために、実証的かつ理論的な証拠を提供する。さらに、適応学習率の分散を補正する項を導入することで、Adamの新しい改良型であるRectified Adam (RAdam)を提案する。画像分類、言語モデリング、ニューラル機械翻訳の実験結果により、我々の直観を検証し、RAdamの有効性と頑健性を実証する。