

# ノイズ注入による平均化を用いた ラージバッチ学習の汎化性能改善手法の検討

長沼大樹<sup>1</sup>  
Hiroki Naganuma

横田理央<sup>1</sup>  
Rio Yokota

東京工業大学 情報理工学院<sup>1</sup>  
School of Computing, Tokyo Institute of Technology

## 1 はじめに

近年、大規模分散深層学習において、大きなバッチサイズでの学習を行うラージバッチ学習では、学習で得られる深層ニューラルネットワークモデルの汎化性能が劣化することが実験的に報告されている [1]。図 1 の黒線で示した学習データのみによる損失関数と緑の破線で示した実社会に存在するデータによる損失関数は必ずしも一致しない。ラージバッチ学習では、局所解から多少のパラメータ変動で誤差が極端に増加する Sharp Minimum に収束することが汎化性能の劣化の原因の一つである [2]。本稿では、ラージバッチ学習における、最適化においても目的関数にノイズを付加し平均化することによる、汎化性能の改善・目的関数の平滑化を検証する。

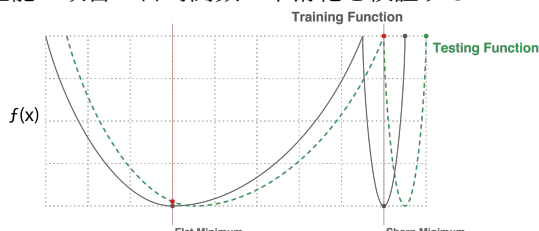


図 1. Sharp Minimum と Flat Minimum

## 2 ノイズ注入による平均化を用いた損失関数の平滑化

Simulated Annealing など古典的な非凸最適化手法では、適切にノイズを与え減衰させることで、大域解への収束を保障する。 $t$  回後の更新を行って得られるパラメータを  $\theta^{(t)} \in \mathbb{R}^d$  として、 $\delta^{(t)}$  をノイズの項、 $\eta$  を学習係数とすると、

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla E(\theta^{(t)}) + \delta^{(t)} \quad (1)$$

のように、学習における更新においてノイズを加え、最適化を行う問題を考える。本稿では、一様分布を用いたノイズ注入を行う SmoothOut [3] をラージバッチ学習に適用し、汎化性能の改善および目的関数の平滑化を検証する。

## 3 実験

ノイズ注入による平均化を用いた損失関数の平滑化実験においては、学習データに前処理を適用せず、バッチサイズを 128 (SB; Small Batch) と 8192 (LB; Large Batch) として、一様分布ノイズを注入した場合のラージバッチ学習の汎化性能改善への影響を測定した。畳み込みニューラルネットワークのモデルとしては VGG9 を用いて学習を行った。Base Line はノイズなしの場合、 $10e-n$  Uniform Noise は一様分布のノイズ  $10^{-n}$  を加えた場合の学習曲線を図 2 に示す。ノイズ注入による平均

化用いない場合に比べ、SB・LB 学習共に汎化性能の改善が見られた。

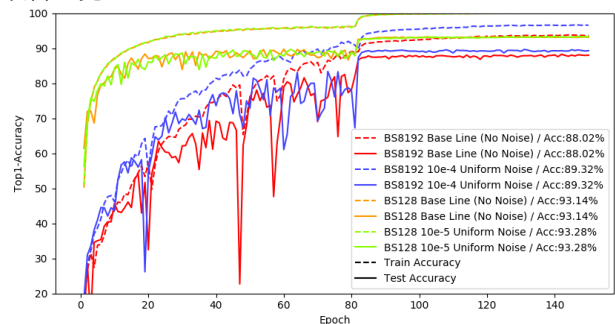


図 2. ノイズ注入による平均化を用いた学習における学習曲線

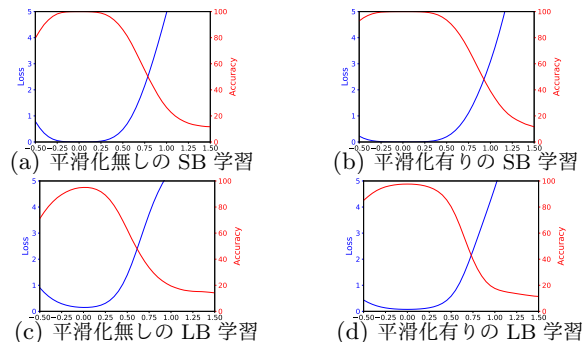


図 3. 学習で得られた解の 1 次元線形補間図。青い線は損失値、赤い線は認識精度を示す。横軸はパラメータ空間の変化量

さらに、損失関数の変化を図 3 の通り可視化した。特にラージバッチ学習では、図 3d の通り損失関数がノイズ注入による平均化により、Sharp Minimum の形状が大きく平滑化していることが確認できた。

## 4 おわりに

今回の実験ではノイズのノルム減衰を行っていないため、Simulated Annealing など古典的な非凸最適化手法としてのノイズ注入による平均化を、ラージバッチ学習においても検証する必要がある。さらに、ラージバッチ学習における汎化性能の劣化と目的関数の平滑性の関係について明確な解明が重要である。

## 参考文献

- [1] Smith S. et al., “A Bayesian Perspective on Generalization and Stochastic Gradient Descent” *ICLR 2018*, 2018.
- [2] Keskar S. et al., “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima”. *ICLR 2017*, 2017.
- [3] Wen W. et al., “SmoothOut: Smoothing Out Sharp Minima to Improve Generalization in Deep Learning”, *arXiv preprint arXiv:1805.07898*, 2018.