

ANNにおける真の最適解による損失ランドスケープの可視化

王 汝沁[†] THAWONMAS Ruck[‡]

立命館大学情報理工学部^{† ‡}

1. はじまり

人工ニューラルネットワーク (ANN), 特に階層を深めた深層学習は他の手法より, 予測精度・汎用性が高いなどの優れた性能を有する. そのため, 多様な視点から, 深層学習モデルの性能解釈が行われ始めている. 本研究では, 特に解空間中における特定な解である「真の最適解」(True Minimum, 以降 TM) の性質探索に着目する.

TMは最高な予測精度・汎用性の解で, 通常の訓練で取得するのが困難である. このため, TMが局所最適解 (LM) または大域解 (GM) より興味深い性質を持つにもかかわらず, 今までの可視化研究では LM または GM の特性分析にとどまっている. 例えば, GMの探索できる勾配法に基づいた改善[1]やLMの損失ランドスケープ可視化[2]に関する研究などが挙げられる.

本論文の目的は, [2]で提出された LM の性質を可視化する方法に基づき, TM の観点から損失ランドスケープの性質を可視化することにある.

2. 提案手法

まず, 表 1 に先述した各解の定義を示す.

カテゴリ	定義
LM	$T_{i+1} \approx T_i$
GM	$T_{i+1} \approx T_i \approx 0$
TM	$T_{i+1} \approx T_i \approx V_i \approx 0$

表 1 解の定義

ここで, T_i は i 回目での訓練損失を示し, V_i は i 回目での検証損失を示す. LM は, 訓練回数の増加は訓練損失にはほぼ影響がない解, GM は訓練損失がほぼゼロになる解と前の研究[1]で定義された. 本論文は訓練損失と検証損失がゼロとなる解を TM と定義する.

次に TM を取得する方法を Algo. 1 に示す. ここで θ を訓練後に得た解, θ^* を TM, x_i, y_i をモデルの入力, 出力とする. LM または GM を見つけたモデル $f(x; \theta)$ に対して, 同モデルの出力 \tilde{y}_i を基準に, 訓練セット X における各データ (x_i, y_i) に新しい教師ラベルを付け直す. その結果を検証用データセット \tilde{X} のデータ (x_i, \tilde{y}_i) とする.

このデータセットに対して, 同じモデルのみで,

Algorithm 1 Get True Minimum

Input: 訓練済みモデル $f(x; \theta)$, 訓練データ $X = \{(x_i, y_i)\}$

Output: \tilde{X}, θ^*

- 1: $\tilde{X} \leftarrow \{\}, \theta^* \leftarrow \theta$
- 2: for i in range($|X|$)
- 3: $\tilde{y}_i \leftarrow f(x_i; \theta^*)$
- 4: Append (x_i, \tilde{y}_i) to \tilde{X}
- 5: return \tilde{X}, θ^*

Algo. 1 TM の取得法

すべての出力は当該教師ラベルと一致することが保証できることになる.

最後に, 損失ランドスケープの計算方法を Algo. 2 に示す. [2]で提出された LM の性質を可視化する方法に基づき, モデル $f(x; \theta)$ に対して, 正規分布より TM の重みベクトル (θ^*) と同じサイズの二つのベクトル δ, η を取る. その後, 損失ランドスケープ S を形成するために, TM の重みベクトルに対して, 変化効果を保つために正規化した $\tilde{\delta}, \tilde{\eta}$ を K 回段階的にインクリメントし, その各変化した重みでの \tilde{X} に対する損失値を記録する.

Algorithm 2 Get Loss Landscape

Input: 損失関数 $L(\tilde{X}; \theta^*)$, 回数 K

Output: 損失ランドスケープ S

- 1: $\delta, \eta \sim N(0, 1)$
- 2: $\tilde{\delta} = \frac{\delta}{\|\delta\|_2}, \tilde{\eta} = \frac{\eta}{\|\eta\|_2}$
- 3: for i in range(K):
- 4: for j in range(K):
- 5: $S[i][j] = L(\tilde{X}; \theta^* + i\tilde{\delta} + j\tilde{\eta})$
- 6: return S

Algo. 2 損失ランドスケープの取得法

3. 実験と結果&考察

3.1 実験

CIFAR-10 公開データセットを使用し, DNN (Dense Neural Network), CNN (Convolution Neural Network), 及び Resnet50 (Residual Network) モデルで TM の損失ランドスケープを可視化した. DNN, CNN, Resnet50 は, 深層学習のさまざまな分野で適用されており, 多数の研究で有効性が証明されているため, 実験対象として選択した.

今回の実験モデルのアキテクチャでは, DNN は 5 層の完全接続層, ノード数が各々 128, 64, 32, 16, 10 である. CNN は各々 128, 64, 32, 16, 10 チャンネルを持ち, コアサイズが 3×3 の畳み込み層を使

Visualization of Loss Landscape from a True Minimum of ANN

[†] Ruqin WANG

[‡] Ruck THAWONMAS*

*ruck@is.ritsume.ac.jp

用する. DNN と CNN は, 最後の層の分類タスクに使用される活性化関数が Softmax で, 他の各層の活性化関数は ReLU である. Resnet50 は 4 つの残差スタックで構成され, それぞれが 3, 4, 6, 3 つの残差ブロックで構成される. 各ブロックには 3 層の畳み込み層とスキップ接続が含まれる. また, パラメータサイズを縮小するためのプーリング層と畳み込み層, および分類のためのプーリング層と完全に接続された層も含む.

CIFAR-10 は分類問題の画像データセットであり, 訓練用の 50,000 枚の写真と検証用の 10,000 枚の写真が含まれている. ベンチマークデータセットとして, CIFAR-10 は深層学習のさまざまなタスクで広く使用されている.

損失関数としてスパースクロスエントロピー (sparse cross entropy), 検証指標として精度を選択する. 最適化手法は Adam, 学習率は 0.01 を設定した.

3.2 ランドスケープの可視化と考察

図 1 の上図, 中図, 下図はそれぞれ DNN, CNN, Resnet50 の損失ランドスケープの等高線と 3D 図を表す.

図 1 から Resnet50 の損失ランドスケープが最も滑らかで, TM 周辺の勾配値が低く, 損失値が下降しやすいため, DNN の損失ランドスケープは非常に鋭く非対称であることがはっきりとわかる.

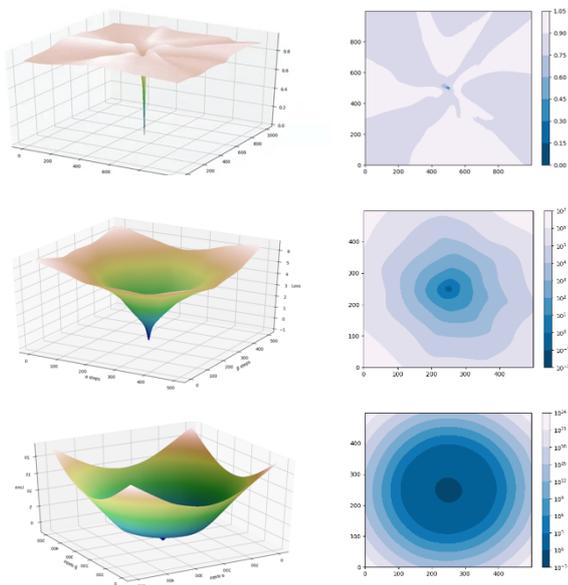


図 1 TM の損失 (上から DNN, CNN, Resnet50)

3.3 練実験と考察

TM を取得するため, [2] の方法に基づき, まずモデルの LM を取得する必要がある. モデルを訓練する時, 幾つかの現象を発見した. CNN と Resnet50 よりも, DNN の LM を取得するための訓練時間が長く, Resnet50 より倍のエポック数が必要であった.

検証損失ほぼゼロになる GM を取得する実験も, DNN は CNN と Resnet50 より倍のエポック数がかかった. ここでも Resnet50 と CNN は効率よく GM を取得できることが分かった. また, Resnet50 は画像のために開発された CNN より必要なエポック数が少ない点も観察した.

図 1 から Resnet50 は損失が徐々に下降し, 値の低い範囲が広い一方で, DNN は損失が急激な下降があり, 値の低い範囲がかなり狭いとわかるため, LM の場合も DNN は勾配法で損失が低い位置に降下するのが更に困難だと推測する. そのため, 実際の訓練では, DNN のパラメータがより低い損失を持つ位置に収束する可能性が低いと示唆された.

4. 終わりに

4.1 まとめ

本実験では, [2] で提案された LM の損失ランドスケープ可視化方法に基づき, TM の損失ランドスケープの可視化アルゴリズムを提案し, 3 種類の ANN のモデルで実験した. 実験に使用したモデル, 生データなどを公開¹している. TM は LM の場合と同じく, Resnet50 は CNN よりも良い性能を持ち, CNN は DNN より性能が良いという結論に至った.

4.2 今後の課題

実際の訓練中に検証損失がほぼゼロになるのは困難の為, モデル訓練の過程で TM に到達できないとも言える. その理由は, 実験結果より TM が周辺より急降下な場所に存在することが示されたため, 最急下降法は TM の位置に到達できないためだと推測する. 訓練中の降下経路を可視化し, 訓練が TM に到達できない理由を観察および分析し, 改善策を模索するのが今後の課題である.

参考文献

- [1] SIMON DU, et al., “Gradient Descent Finds Global Minima of Deep Neural Networks,” PMLR 97:1675–1685, 2019.
- [2] HAO LI, et al., “Visualizing the loss landscape of neural nets,” NIPS 2018 (extended version, 10.5 pages). <https://github.com/tomgoldstein/loss-landscape>.

¹ <https://tinyurl.com/wt83nc-ipsj>