

LeNet-5 における学習ネットワークの分析

山城海人[†] 廣津登志夫[‡]

[†]法政大学情報科学部

1 序論

近年, Artificial Neural Network(ANN) が注目され様々な分野に利用されている. ニューロンを最小単位として, ニューロンからニューロンへ信号が伝達される. このニューロン間の信号には固有の重み付けがされており, 学習中に重みを変更することで ANN は学習する. ANN は解決したい問題に応じて, 時系列データには Recurrent Neural Network(RNN) が, 画像認識には Convolutional Neural Network(CNN) が広く使われている. ANN は各種パラメータを適切に調整することで, 各問題に対して最適化を行う. そこで, すでに精度が認められている CNN, LeNet-5 のネットワークを調査し, 初期配置や学習データのデータセットが CNN の精度や成長にどの程度影響を与えるのかを示す.

2 LeNet-5 の概要

LeNet-5[1] は YANN LeCun によって提案された CNN である. CNN が学習する際には誤差関数を用いて, 正解のラベルと CNN が出力した結果の差分を求める. 今回利用した交差エントロピー誤差関数は以下の式で表される.

$$E = - \sum_k t_k \log e y_k$$

y_k は CNN の出力を示し, t_k は学習データの正解ラベルを示す. t_k は one-hot 表現で表される. y_k には各ラベルについて CNN から出力された値が格納されている. 正解ラベルに対する y_k の値が大きいほど交差エントロピー誤差の値は 0 に近づいていく. 0 に近づくほど CNN が正しい答えを出せていることを意味する. CNN は交差エントロピー誤差の値が 0 に近づくように重みを変更する. 重みを変更する際に, 一度に重みを変更すると CNN の学習が収束しない可能性がある. そこで学習率を設定する. 用いた学習率は以下の式で

表され表 1 はパラメータを示す.

$$base_lr * (1 + gamma * iter)^{-power}$$

base_lr	momentum	weight_decay	gamma	power
0.001	0.9	0.0005	0.0001	0.75

表 1: 使用したパラメータ一覧

重みを更新する手法として今回は Momentum を利用した. Momentum は次の式で表される.

$$v = \alpha v - \eta \frac{\partial f}{\partial x_0}$$

$$x_0 = x + v$$

η は学習率, x は各ニューロンの重みの値を示す. v は速度を表し初期値は与えられず重みを更新する際に重みと同じ値が与えられる. α を Momentum といい予め 0.9 という値を与えた. 重みの初期化には Xavier[2] を利用している. Xavier アルゴリズムは以下の式で表される分散を持つ正規乱数を生成する.

$$Var = \frac{2}{n_{in} + n_{out}}$$

3 実験

データセットと重みの初期値の影響を調べるために, データセットは Chars74k と MNIST の 2 種類を用いる. 初期値は Xavier を用いて決定する. MNIST が数字データのためのため, Chars74k の数字データ以外のものは用いない. また, Chars74k と MNIST は白黒値が反転した画像となっているが実験上, 色が反転していると精度が出なかったため, MNIST の白黒値を反転することで Chars74k と揃えた. Chars74k, MNIST それぞれ 5 つのモデルを用意した. それぞれのモデルを Chars1~5, MNIST1~5 とする. Chars1~5 は重みの初期値が異なる. Chars と MNIST の数字が対応するモデルは重みの初期配置が同じである. Chars1~5 と MNIST1~5 のそれぞれ比較することで重みの初期配置が異なるが, 学習に使うデータセットが同じ場合の比較ができる. Chars と MNIST の数字が対応するモ

Analysis of Learning network on LeNet-5

Kaito YAMASHIRO[†], Toshio HIROTSU[‡]

[†]Faculty of Computer and Information Sciences, University

	Chars1	Chars2	Chars3	Chars4	Chars5
Chars1	0.0	841.5	849.4	836.2	807.3
Chars2	841.5	0.0	858.4	824.9	840.9
Chars3	849.4	858.4	0.0	820.3	824.8
Chars4	836.2	824.9	820.3	0.0	815.8
Chars5	807.3	840.9	824.8	815.8	0.0

表 2: 重みの初期値が異なり Chars74k を使って学習した LeNet-5 の重みの差分和

1	2	3	4	5
200.5	203.0	225.7	208.7	219.8

表 3: 重みの初期値が同じ Chars, MNIST を使って学習した LeNet-5 の重みの差分和

デル同士を比較することで重みの初期配置が同じだが、学習に使うデータセットが異なる場合の比較を行う。

3.1 重みの値の差

モデルの重みの値がどれだけ違うのか検証する。用意したモデルはすべて同じ LeNet-5 の構造をしているため、重みの値を持っている変数の数が同じである。よってモデル間で同じ位置の重みの差分のとり、モデル全体で差分の絶対値和を取ることで2つのモデル間での重みの構造の差を示す指標とする。表2は Chars1~5 のモデル同士で差分を求めたものである。Chars1~5 と MNIST1~5 の数字が対応するモデル同士でも同様に差分をとった結果が表3である。比較するとわかる通り、重みの初期値が違う場合よりも重みの初期値が同じ場合のほうが差分和の値が小さい。

3.2 学習の損失関数の値

CNN の精度をはかる指標の一つに損失関数がある。損失関数の値が0に近いほど CNN は理想的な解をだすことができる。図1は Chars74k の15を1000回学習させた時の損失関数の値の推移である。横軸は何回目の学習かを表し、縦軸が損失関数の値を示す。重みの初期配置が違い、構造の違いが大きいにもかかわらず5つのモデルとも似たような学習経過をたどる。この結果は Chars74k だけでなく、MNIST で学習させたモデルも同様の結果が見られた。図2は Chars1 と MNIST1 の損失関数の値の推移をグラフにしたものである。図1のデータセットが同じモデル同士のグラフと比べると、2つとも違う経過をたどっている。

4 まとめ

精度や損失関数の値の推移は、初期配置が異なり最終的なネットワーク構造が異なる場合でも、学習デー

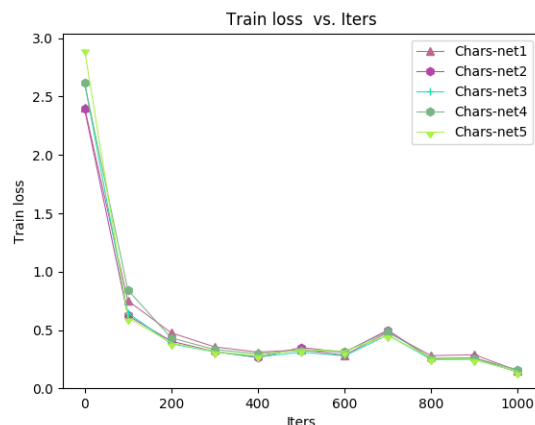


図 1: Chars で学習したネットワークの損失関数の値

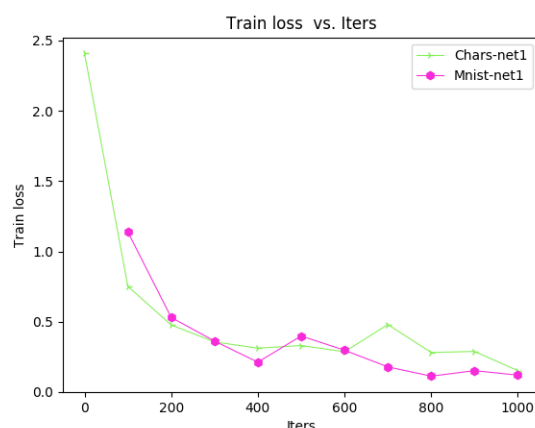


図 2: 重みの初期値が同じでデータセットが違う損失の変化

タの影響が大きいことがわかった。

謝辞

本研究は JSPS 科研費 JP17K00138 の助成を受けたものである。

参考文献

- [1] Y. LECUN, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 9, 2010, pp. 249–256.