

動的隠れ層を考慮した多層ニューラルネットワーク

申請中 法政大学 *天海透 AMAGAI Toru
01309090 法政大学 千葉英史 CHIBA Eishi
法政大学 東原正智 HIGASHIHARA Masanori

1. はじめに

隠れ層を多層化することによって汎化能力が高まり、ニューラルネットワークはここ数年で急激に発展した。一方で、計算の量が膨れ上がったため、学習に大変な時間がかかるようになった。

既存の多層ニューラルネットワークでは、隠れ層の個数は経験に基づいて定められる。本研究では、動的に隠れ層を生成しながら（隠れ層の個数を増やししながら）学習を進めるニューラルネットワークを扱う。計算実験から、提案したニューラルネットワークが従来のニューラルネットワークと比較して、正答率を大幅に改善するだけでなく学習時間を減らす効果があることを確認した。

2. 勾配消失

隠れ層を増やしすぎることによって学習がうまく進まないことがある。その理由として挙げられるのが勾配消失である。

出力層からの逆伝播において、入力層側のニューロンに対する勾配が非常に小さいとき、重みの更新量がほぼゼロになる。この現象が勾配消失である。層を逆伝播する毎に、誤差は小さくなるので、層を増やすほど勾配消失が起りやすくなることが知られている。

3. 提案手法

前章で述べたように、勾配消失の起りやすさは、隠れ層の個数に依存する。そこで本研究では、動的に隠れ層の個数を増やししながら学習を進める手法を提案する。具体的には、隠れ層の個数が少ない状態から始め、Lossの差に関する条件を満たすたびに、出力層側の隠れ層の手前に新たな隠れ層を追加する。

文献 [1] では、動的にニューロン数を増やししながら学習を進める手法に注目した。文献 [1] の多層ニューラルネットワークでは、隠れ層の個数は変化しないため、汎化能力向上に関する保障はない。本研究の提案手法では、徐々に隠れ層を増やすた

め、汎化能力が高くなる。隠れ層を増やすときの流れを図1に示す。

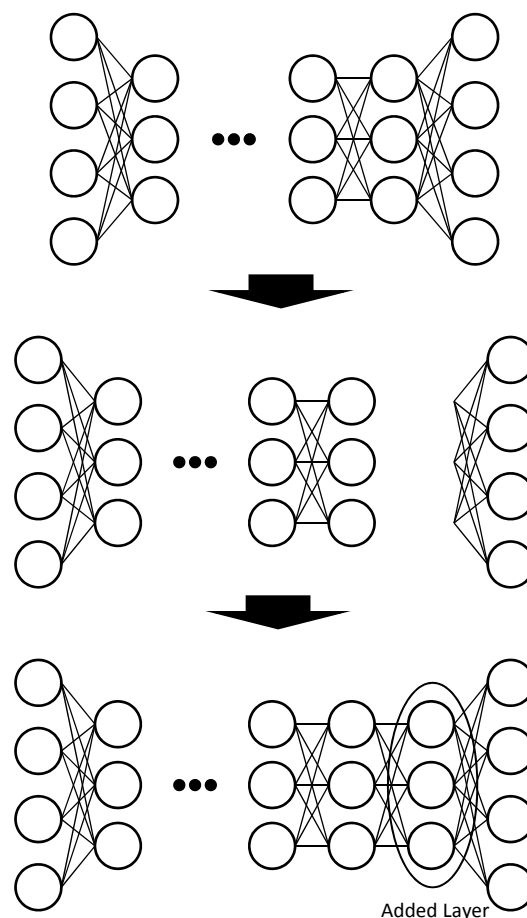


図 1: 隠れ層を増やすとき。

隠れ層を増やすときの条件は、誤差関数によって求められる Loss に基づく。具体的には、学習の前後で Loss の差が与えられた定数以下という条件である。隠れ層を増やすと勾配消失が起りやすくなる状況になるが、提案手法では入力側のニューロンに対する事前学習が、既にある程度進んでいることから、勾配消失の影響を受けにくい。隠れ層を増やす位置に関して、出力層は勾配消失の影響

を受けないので、出力層側に新たな隠れ層を追加する。

提案手法は、従来のニューラルネットワークと比較して、隠れ層が少ない状態から学習を開始するため、学習するニューロンの個数が少ない。そのため、提案手法における計算の量は、従来のニューラルネットワークと比べて相当少ない。

4. 実験結果

0~9の手書き数字の画像をクラスタリングするためのデータセットを用いる。初期設定において、全ての重みパラメータは平均0分散0.1の正規分布に従い、隠れ層の数を1とする。各隠れ層中のニューロン数は250とする。隠れ層と出力層の活性化関数として、それぞれReLU関数とsoftmax関数を使う。最適化手法にSGDを使う。総Epoch数は500にする。1Epoch毎に128枚の画像を学習する。

Cross_entropy_error関数から得られるLossに関して、Epochの前後でLossの差が0.0025以下のとき、隠れ層を1つ増やすとする。また、隠れ層の個数は18層まで増加するとする。正答率に関する実験結果を図2に示す。さらに、Epoch数と層数の関係、および、各隠れ層数に対する最大正答率を表1に示す。

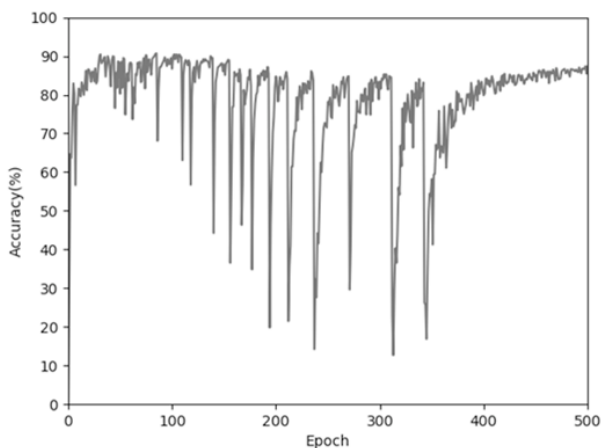


図 2: 実験結果

隠れ層の数を18に固定したニューラルネットワークでは、500Epochを通して、正答率が10%前後を示した。一方、提案手法では、各隠れ層の数に対する最大正答率は非常に高い。隠れ層数を固定

表 1: 実験で得られた Epoch 数, 隠れ層の数, 最大正答率の関係。

Epoch 数	隠れ層の数	最大正答率
0~6	1	82.92
7~44	2	90.45
45~54	3	89.29
55~61	4	88.32
62~85	5	90.69
86~109	6	90.29
110~117	7	88.91
118~139	8	89.96
140~155	9	88.86
156~166	10	86.58
167~175	11	86.52
176~193	12	86.27
194~210	13	85.90
211~236	14	86.11
237~270	15	84.55
271~311	16	85.29
312~342	17	84.05
343~500	18	87.03

したニューラルネットワークにて、正答率が低い原因は、勾配消失のためと考えられる。提案手法では、隠れ層数が18のときでも、勾配消失の影響を受けにくく、正答率が87%を超えた。また、提案手法では、隠れ層数が5のときの最大正答率が、他の隠れ層数のときと比べて最も高い(90.69%)。この結果は、隠れ層の数を増やす条件が改善できることを示唆する。

5. 今後の課題

今後の方針として、多層ニューラルネットワークの構造について、全結合層だけでなく、畳み込み層も考えられる。隠れ層の数を増やす条件に関する考察が今後の課題として考えられる。

参考文献

- [1] M.R. Azimi-Sadjadi, S. Sheedvash, and F.O. Trujillo, "Recursive dynamic node creation in multilayer neural networks," IEEE Transactions on Neural Networks, vol.4, no.2, pp.242-256, 1993.