ISSN 2186-5647

2-28

スパース性を用いた多層ニューラルネットワークのリンク最適化

日大生産工　○田村 匡　　日大生産工　山内　ゆかり

1. まえがき

機械学習の研究は盛んに行われており、画像分類や言語処理など、様々なタスクに使用されている。機械学習の一種であるニューラルネットワークは、ノードとエッジで構成されたネットワークで、入力層から出力層までの間の隠れ層の数とノード数によって、性能が変化する。総数やノード数が大きいほうが表現能力を向上させるものの、学習にかかる時間が伸び汎化性能を低下させる可能性もあるため、問題毎にパラメータチューニングをする必要があった。

2. 従来手法

階層型ニューラルネットワークは、誤差逆伝搬法[1]により、出力値と期待する出力の誤差を重みの関数とすることで、勾配法等を用いて重みを更新して誤差を減少させる学習モデルである。このモデルは回帰やクラスタリング問題を解くことができるが、問題の種類や複雑さによって適切な層数や各層のノード数を事前に決定する必要がある。佐々木らは層数を動的に最適化するために動的多層化ニューラルネットワーク[2]を提案し、少ない層数から徐々に層数を増やして最適な層数を決定して学習することで認識率を向上させた。しかし、メモリと計算時間が膨大で、複雑な問題に対しても層数があまり増えず表現力も十分でなかった。

本研究では、層数ではなく各層毎のリンクをスパース性を用いて最適化することで、認識率を向上させると同時に計算コストを低下させる手法を提案する。

3. 提案手法

提案手法ではスパース性を取り入れて重要性の低い重みを0に近づけることを考える。誤差関数にL2ノルム項を追加することで、降下法を用いて各重みを0に近づく方向にも移動させる。

$$\mathrm{E} = \mathrm{E(z)} + \frac{\lambda}{2}\sum w^2 \quad (1)$$

$$\frac{\partial \mathrm{E}}{\partial \mathrm{w_i}} = \frac{\partial \mathrm{E(z)}}{\partial \mathrm{w_i}} + w_i \quad (2)$$

誤差関数E(z)は問題によって変わるが、これにすべての重みの2乗和を足したものを誤差関数とする。これを適応させることで、過剰にあるノードと重みのうち重要度の低いものの値がほぼ0となり、モデルへの影響度を小さくさせることができると考えられる。

4. 実験方法及び測定方法

モデルの性能評価のために、MNISTをデータセットとして用いたクラス分類問題を解く。隠れ層が2層の誤差逆伝搬ニューラルネットワーク用いて、学習データのうち3200個のデータを抽出して1エポックとし、50エポックの学習を10回行い、認識率、学習データに対する誤差関数値、テストデータに対する誤差関数値、ノルムの4つについて計測した。

クラス分類問題なので出力層の活性化にはsoftmax関数を用い、誤差関数は交差エントロピーを用いた。

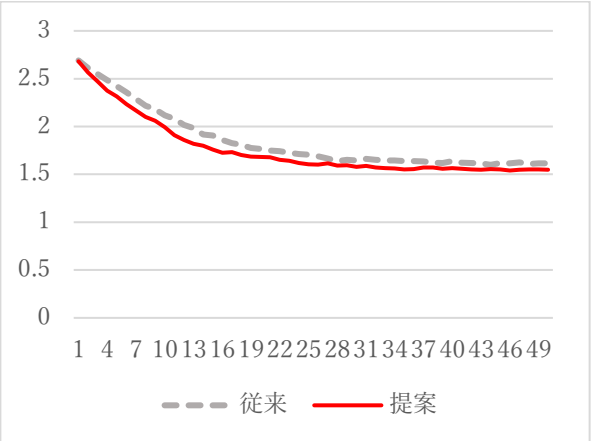
$$\mathrm{E(z)} = -\sum t * \log(z) \quad (3)$$

実験に使用したPCのスペックはWindows10 Pro, CPU Intel Corei7-8700, RAM 16GB, GPU GeForce GTX1080 である。

5. 実験結果及び検討

スパース項の係数λが0.1の場合の、交差エントロピーをFig1. Fig2.、文字認識率をFig3.、L2ノルムをFig4.に示す。またTable1.に係数λ変更時の各値を示す。

Fig1. 学習データでの誤差



Sparse Representation of Dynamic Multilayer Neural Networks
Masashi TAMURA and Yukari YAMAUCHI

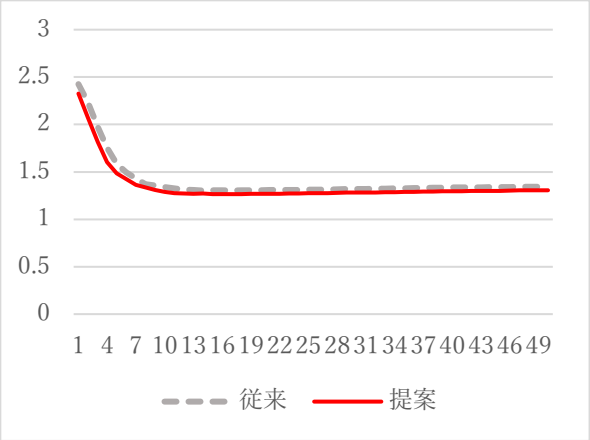


Fig2. テストデータでの誤差

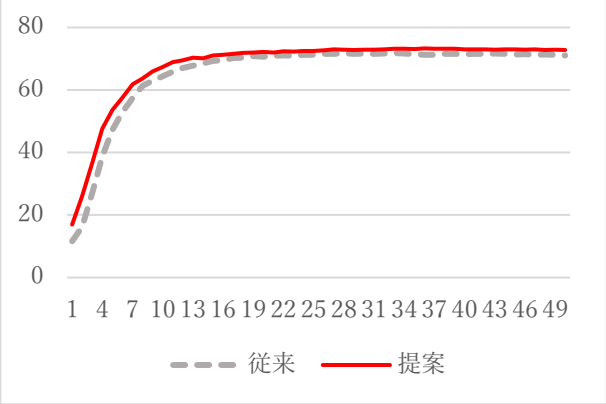


Fig3. 認識率

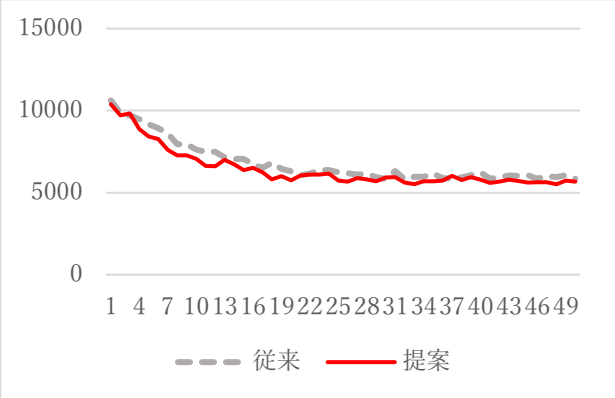


Fig4. L2ノルム

Table1. 係数λ変更時の各値

係数	学習時誤差	テスト時誤差	認識率	L2 ノルム
0	1.615	1.348	71.03	5855.834
0.02	1.590	1.338	71.92	5669.863
0.1	1.548	1.307	72.75	5669.079

表1より、L2ノルムの減少に伴って、学習時及びテスト時の交差エントロピーも低下したことがわかる。向上した理由としては、学習初期は解のスパース推定により収束が早まり、学習終盤では重みのノルムを低下させることで汎化性能を損なわないようになったことがあげられる。

6. まとめ

本論文ではニューラルネットワークのリンク最適化を行った。誤差関数にL2ノルムの項を追加してリンクのスパース推定ををすることで、認識率の向上を実現できた。

[1] D.E. Rumelhart, G.E. Hinton, and R.J. Williams,
"Learning representations by back-propagating errors,"
Nature, vol.323, pp.533–536, 1986.
[2] 佐々木 駿也, 萩原 将文 動的多層化ニューラルネットワーク, 電子情報通信学会論文誌 D Vol.J102-D No.3 pp.226-234