

CNNにおける非線形活性化関数の構成とその適用

日大生産工(学部) ○榎本 知幸

日大生産工 角田 和彦

日大生産工 三浦 慎一郎

1 まえがき

近年では、畳み込みニューラルネットワーク (CNN) [1]を用いた人工知能の画像認識能力の精度は高く、PET/CTの画像情報による、肺結節の自動検出[2]など、医療の分野における活躍も期待されている。

そのCNNの認識能力を高める手法の一つに、活性化関数の最適化が挙げられる。CNNにおける畳み込み層の活性化関数として、現在、ReLU関数[3]が幅広く使われているが、PReLU関数[4]やELU関数[5]など新しい活性化関数が提案されており、どの活性化関数がどのネットワークにおいて最適であるかは分かっていないことが多い。

本研究では、流体力学的立場から得られる非線形活性化関数[6]をReLU関数に組み込み、PReLU関数のようなパラメータを持つ活性化関数を提案する。また、その活性化関数を、画像認識を行うCNNに適用することで、パラメータを持つ活性化関数の特性を明らかにすることを目的としている。

2 畳み込み層における活性化関数

CNNにおける畳み込み層の活性化関数として、シグモイド関数、tanh関数など多くの関数が使われてきたが、現在ではReLU関数が幅広く使われている。この関数は入力値 v が入力されると、 $\max(0, v)$ を出力値として返す。

ReLU関数は勾配消失問題を軽減することが出来るという特徴があり、畳み込み層の活性化関数に適用することでシグモイド関数やtanh関数よりも良い結果が得られている。

また、ReLU関数の負の領域を拡張したPReLU関数はReLU関数よりも優れた結果を残した例も報告されている。

PReLU関数は次のように示される。

$$f(v) = \begin{cases} v & (v > 0) \\ av & (v \leq 0) \end{cases} \quad (1)$$

ただし、パラメータ a は逆伝搬により学習し、活性化関数は最適化される。

3 流体力学的立場から構成されるReLU関数の近似関数

文献[6]では次のような活性化関数が示されている。

$$f(v) = \begin{cases} \frac{1}{2} \left(2 - \frac{1}{1+|\gamma|} \right) & (v > 0) \\ \frac{1}{2} \left(\frac{1}{1+|\gamma|} \right) & (v \leq 0) \end{cases} \quad (2)$$

ただし、 $\gamma = v/2k$, $k > 0$ 。

この活性化関数では、ネットワーク温度 k を用いて、ステップ関数からシグモイド関数を近似することができる。

そこで、負の入力に対する式に注目し、入力値 v をかけ、係数を除去することで次の式が得られる。

$$g(v) = \frac{kv}{k-v} \quad (v < 0) \quad (3)$$

この式は $k \rightarrow 0$ の極限で傾き 0、 $k \rightarrow +\infty$ の極限で傾き 1 の線形関数に収束する。これをReLU関数の負の入力における式に適用し、ネットワーク温度 k を、学習パラメータ a を含む $\exp(a)$ に置換することで次の活性化関数が得られる。

$$h(v) = \begin{cases} v & (v \geq 0) \\ \frac{e^a v}{e^a - v} & (v < 0) \end{cases} \quad (4)$$

この活性化関数は $a \rightarrow -\infty$ の極限でReLU関数に収束し、 $a \rightarrow +\infty$ の極限で傾き 1 の線形関数に収束する (図1)。

また、逆伝搬学習のために使われる v の偏微分は次式で示される。

Construction of Nonlinear Activation Function in CNN and Its Application

Tomoyuki ENOMOTO, Kazuhiko KAKUDA and Shinichiro MIURA

$$\frac{\partial h(v)}{\partial v} = \begin{cases} 1 & (v \geq 0) \\ \left(\frac{e^a}{e^a - v}\right)^2 & (v < 0) \end{cases} \quad (5)$$

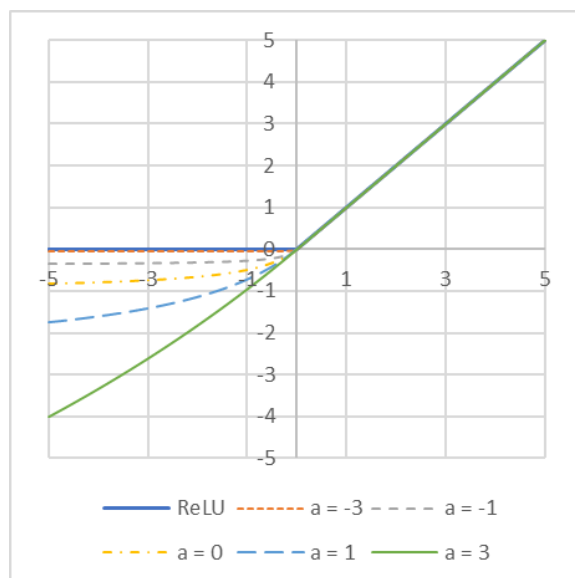


図1. 式(4)の活性化関数

また、活性化関数における a による偏微分は次のように示される。

$$\frac{\partial h(v)}{\partial a} = \begin{cases} 0 & (v \geq 0) \\ -\frac{e^a v^2}{(e^a - v)^2} & (v < 0) \end{cases} \quad (6)$$

しかし、この式によって得られる勾配は非常に緩やかなため、学習時は定数 β を掛けることで学習を高速化する。

CNNへの実装時は、各カーネルで共有するパラメータ a_i を導入する。

それに従い、誤差関数 E によりパラメータ a_i は、次のように更新される。

$$\Delta a_i := \mu \Delta a + \beta \varepsilon \frac{\partial E}{\partial a_i} \quad (7)$$

$$\frac{\partial E}{\partial a_i} = \sum_{v_i} \frac{\partial E}{\partial h(v_i)} \frac{\partial h(v_i)}{\partial a_i} \quad (8)$$

ただし、 μ はモメンタム係数、 ε は学習率、 $\sum v_i$ は a_i に対応する全ての v_i における合計を表す。

4 MNISTでの数値計算

ReLU関数と式(4)の活性化関数をCNNの畳み込み層に組み込み、手書き数字データセットMNISTを使ってそれぞれ3回学習させた。実験に用いる畳み込み層の構造を表1に示す。ま

た、畳み込み層の次の層にはプーリング層を入れている。

各ハイパーパラメータは $\mu=0.9$ 、 $\beta=10000$ 、 $\varepsilon=0.001$ に設定し、全ての a_i の初期値は0.0とした。表2は、それぞれの関数の結果の最高値と最低値を表している。

表1 実験に用いるCNNの構造

フィルタサイズ	5・5
カーネル数	16・32

表2 実験結果

	ReLU	本研究の活性化関数
最高正解率	0.9784	0.9800
最低正解率	0.9721	0.9753

5 まとめ

本研究では、ネットワーク温度による関数の分布の変化といった特徴を持つ、流体力学的立場から構成される活性化関数を、PReLU関数に組み込み、新しい活性化関数を提示した。

MNISTでの実験では、3回と少ない試行回数ではあるが、この実験で用いたCNNに関しては、この活性化関数はReLU関数と同等か、それ以上の学習能力を持っているのではないかと考えられる。

今後は一般物体認識のデータセットや、より大きなCNNで実験し、この活性化関数を使ったCNNの性能について詳しく調べていきたい。

「参考文献」

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proc. of the IEEE, (1998), pp.2278-2324.
- [2] 寺元篤司, 藤田広志, 山室修, 玉木恒男, "PET/CT画像におけるConvolutional neural network と従来型の特徴量を併用した肺結節自動検出手法の改良", 信学技報 vol116 no.39 MI2016-4, (2016), pp17-20
- [3] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines", In ICML, (2010), pp.807-814.
- [4] K. He, X. Zhang, S. Ren, J. Sun, "Delving Deep into Rectifiers", arXiv:1502.01852v1 [cs.CV] 6 Feb 2015.
- [5] D. Clevert, T. Unterthiner, S. Hochreiter, "Fast And Accurate Deep Network Learning by Exponential Linear Units (ELUs)", arXiv:1511.07289v5 [cs.LG] 22 Feb 2016.
- [6] 角田和彦, "ニューラルネットの特性関数とその近似関数", 情報処理学会第67回全国大会, (2005), pp.239-240.