

Neurons regarded as Graph

江嶋 宏優

Atsuhiro Ejima

熊本大学 理学部

Faculty of Science, Kumamoto University

Deep learning has revolutionized various fields by emulating brain structures found in living organisms. While mathematical approaches are fundamental in studying machine learning, this paper argues for the essential role of a biological perspective. A novel layer closely mimicking neuron structure is proposed, and its performance is compared against existing algorithms. The proposed layer exhibits distinctive characteristics when compared to existing algorithms, highlighting its potential impact on machine learning applications. It is observed that increasing the number of recurrences typically leads to a loss of features in neural networks without optimization. However, the results of this paper differ from them and might show the effectiveness of recursive processing. Overall, this study emphasizes the significance of considering biological inspiration in deep learning models and the potential benefits of the proposed layer for improving the performance in machine learning tasks.

1. はじめに

ニューラルネットワークは、脳の思考プロセスを模した数理モデルであり、一般的な手法として誤差逆伝播法によるパラメータの調整が行われる。多層ペーセptronが発明され、それを元にして数多くのモデルが提案された。畠み込み処理や再帰型のアルゴリズムは画像認識や自然言語処理において驚くべき功績を残した。しかしながら、脳の解明は未だ完全ではなく、推測による理論の構築では生物の思考プロセスを忠実に再現が難しい。例えば、全結合層はニューロンが全ての結合を有する点で実際の構造と異なる。加えて、神経細胞間で連鎖律による偏微分での誤差伝播が計算されているとは考えづらい。深層学習の基礎を築いたGeoffrey Hinton氏が2022年に公開した論文「The Forward-Forward Algorithm: Some Preliminary Investigations」で述べられたように、大脳皮質は誤差を明示的に伝播させたり、逆伝播処理のために神経活動を保存したりするという確証的な証拠は存在しないと考えられている。

ただし、分類や回帰、クラスタリングといった機械学習タスクにおいてニューロン構造を厳密に再現することは非効率な可能性があり、必要性について言及しなければならない。確かに分類では指定されたクラスに属する確率をソフトマックス関数により求め、最も高い値が出力とされるが、これは生物の脳構造とは大きく異なる。MNISTデータでの画像分類を例にすると、事前に0~9のクラスに対応するニューロンを用意すること自体がまるで事前にその存在を知っていたかのように感じられる。ヒトは元々数字の概念を知らず、その存在のためのニューロンが脳内に存在しているわけではない。このように、機械学習と実際の神経細胞が構成するネットワークは異なるものの、それを模して数理モデルを開発した、幅広い問題を解けるようにすることは非常に有効である。

本論文では、生物の思考能力を全結合層よりも具体的に再現するため、局所的な結合を持ち、恒常的な値の送受が起こるニューラルネットワークの層を考察し、その意義や問題点について考察する。

2. グラフとしてのニューロン

多層ペーセptronが生物の脳を模して発明されたように、より鮮明にそれを模擬することが機械学習の発展には必要かもしれない。生物の脳では、物理的に近接した神経細胞同士により、局所的な結合が形成される。具体的に、ニューロンは細胞体から軸索を通じ、シナプスにて神経伝達物質が運ばれることにより、電気信号のやり取りを行う。このことから、ニューラルネットワークは特に有向グラフとして一般化できる。

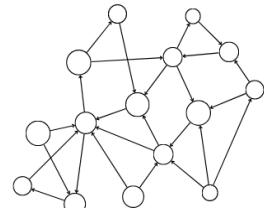


図 1: 結合の状態を有向グラフで示した例

その他の再現方法としては、各ユニットに座標を与え、そのユークリッド距離を計算し、接続距離を指定した確率で与えることができる。例として、複数のニューロンに対してランダムな座標を与え、接続距離をdとして局所的な結合を再現する。

$$f(x) = -\frac{1}{d}x + 1$$

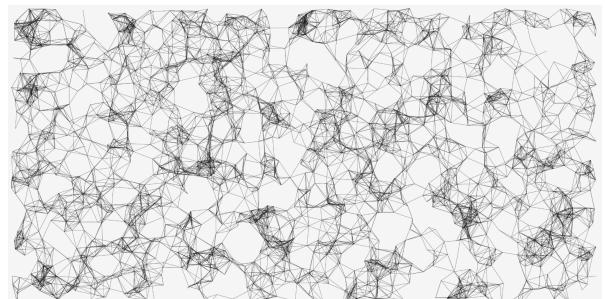


図 2: 座標による結合の再現
ニューロン数2000, d=0.8, shape=[20, 10]

図2にニューロン数2000、接続距離0.8、ネットワークの形状[20, 10](2次元)での接続例を示した。点(0, 1), (d, 0)を通る一次直線にて接続の確率密度関数を定義すると、上記式のように表され、実際に実行すると次のような結合状態が生成される。このように、座標を用いた生成も可能ではあるが、本質的にグラフ構造と差異のないことや実装上の問題点などから、本論文ではグラフ構造としてニューロンを表現する。

2.1 Graph Layer

機械学習の歴史において、畠み込みニューラルネットワーク(CNN: Convolutional Neural Network)や再帰型ニューラルネットワーク(RNN: Recurrent Neural Network)は従来のアルゴリズムと比較して、精度を大幅に向上させたことは良く知られている。それに続く新たなアプローチとして、グラフニューラルネットワーク(GNN: Graph Neural Network)が考案された。GNNは人の繋がりや地点の関係性を表すことができ、実際にいくつかの手法が考案されてきた。

今回提案するGraph Layerは、GNNとは生物のより鮮明な再現を目的としている点で異なる。そして、グラフ構造に基づいた値の送受により、局所的・再帰的な伝播が可能なためCNNとRNNの2つの特徴を組み合わせた層だと見なされる。

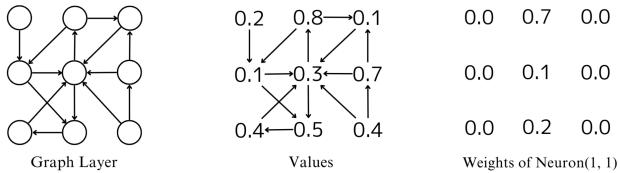


図3: ニューロン(1, 1)における結合状態

図3で示されるように、接続距離を1とした場合の座標が(1, 1)のユニットにおける接続配列W[1][1]は要素に接続度合いを保存する。Graph Layerは従来のネットワーク構造に組み込んだ利用を想定するため、一般的な入力・出力の機構に対応する。しかし、局所的な値の伝播はGeoffrey Hinton氏が提唱したForward-Forward Algorithmと似た性質が期待できるかもしれない。

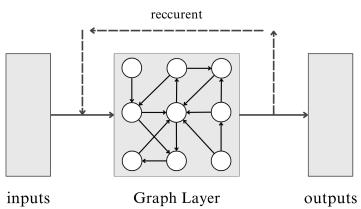


図4: Calculation Flow

図4に示されるように、Graph Layerは指定された回数の反復によって値の送受を行うため、再帰的な計算が実現し、接続距離を超える範囲にも影響を及ぼすことが特徴的である。例として、入力が28×28画像の2次元配列を想定した場合、Graph Layerの形状も同じく28×28の2次元テンソルになり、各ユニットは接続距離に応じた重み配列を持つ。実装では入力が(バッチサイズ, 784)であることを想定し、Graph Layer内で2次元形状への変換が行われる。また、接続距離を超える範囲への値の送受は不可能であるが、これはゼロパディングや範囲を指定した計算により回避できる。反復回数を1に設定し、パディングを行わない場合は図5の式を指定された回数繰り返した結果が

outputとなり、各ユニットが持つ重みであるweight[y][x]の形状は、接続距離をdとすると、縦横の長さは(2d+1)である。

$$output[Y][X] = \sum_{y=d}^{27-d} \sum_{x=d}^{27-d} input[y][x] * weight[y][x]$$

$$Y, X = slice(y - d, y + d), slice(x - d, x + d)$$

図5: loop=1, without zero padding

2.2 問題点

まず、出力の決定に関する問題点について言及する。従来の手法では、全結合層などを使用することで出力の形状を指定し、誤差関数による検証が可能であった。しかし、Graph Layerでは入力と出力の形状が維持され、one-hotエンコードされたベクトルに変換することができない。そのため、単体での出力の決定が困難となる。生物の思考プロセスの観点からすると、出力を1つの値に絞り込むことに違和感を覚えることがある一方で、現在の機械学習タスクでは、予測値を明確に決定する必要がある。よって、モデルはアフィン変換などを使用して、出力を決定する。

次に、パラメータの合計数に関する問題として、メモリの使用量が膨大になり得ることが考えられる。畠み込み処理では、各ピクセルに対してカーネルを利用して計算することで、少ないパラメータで演算を行うことができる。Graph Layerではパラメータ次第では膨大で処理の困難な演算量になりえるため、注意が必要である。生物のニューロンにパラメータ数に関する制約は考えづらいが、膨大な計算資源が求められる場合には全結合層や次元削減によって、入力サイズを小さくし、パラメータ数を減らすことが有効である。

最後に、特有の値の送受方法による問題点を指摘する。確かにニューロンは物理的な結合を持つために局所的な値の送受は意味のあることかもしれないが、大域的な影響を受けづらく、パラメータ調節によっては精度が低下することが予測される。これに対する解決策として、適切な局所的結合を可能にするため、全結合層による値の変換を行うことが有効である。また、過剰な再帰回数の場合では、入力の特微量が失われてしまう可能性も考えられる。図6に示されるAとBは重みが共有され、ランダムなrgb値を入力とした場合を30再帰回数単位ごとに可視化した例である。

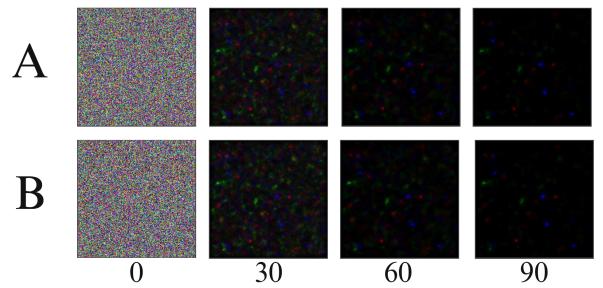


図6: 再帰回数による特微量の消失

異なる入力にも関わらず、計算処理が続くことにより非常に似た出力結果が得られる。これは重みの共有による偏りが原因であり、本来の情報が失われるかもしれない。しかしながら、この回路が欠点であるかは実験によって確かめることが重要である。そのため、再帰処理における重みの更新効果や局所的な計算が精度に影響を与えるのかを検証する。

3. 実験

Graph Layerを実装し、画像分類タスクにおいてその精度や効果を検証する。実験には、一般的に機械学習モデルの評価によく使用されるデータセットであるMNISTとFashion-MNISTを使用した。どちらも10クラスで 28×28 ピクセルの画像として提供されている。本実験での主な目的は全結合層との比較です。近年の機械学習アルゴリズムにはAttention機構やTransformerモデルなどがあり、非常に高精度な結果をもたらすが、これらには人為的目的に対する手段が含まれると考えられるため、基礎技術である多層パーセプトロンを比較対象とした。

3.1 モデル構造と最適化手法

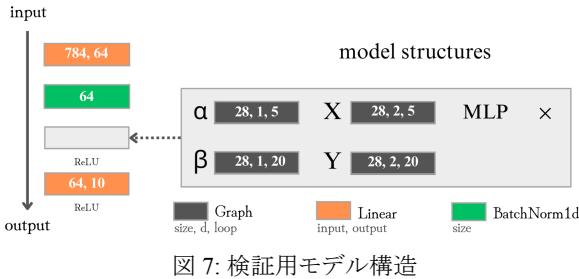


図 7: 検証用モデル構造

接続距離および再帰回数による影響や既存のアルゴリズムとの比較を行うために、基本的な構造を共通に持つ複数のモデルを検証する。図7に示すように、1つのMLP(Multilayer perceptron)と、4つのGraph Layerを使用したモデルを比較する。これらのモデルには全結合層とバッチ正規化のみが利用されている。最適化手法は学習率を0.001とし、勾配の一次モーメントと二次モーメントを利用して、パラメータごとに学習率を決定する最適化手法であるAdamを用いた。

$$\begin{aligned} \mathbf{m}_{t+1} &= \beta_1 \mathbf{m}_t + (1 - \beta_1) \frac{\partial L}{\partial \mathbf{W}} \\ \mathbf{v}_{t+1} &= \beta_2 \mathbf{m}_t + (1 - \beta_2) \frac{\partial L}{\partial \mathbf{W}} \frac{\partial L}{\partial \mathbf{W}} \\ \hat{\mathbf{m}}_{t+1} &= \frac{\mathbf{m}_{t+1}}{1 - \beta_1^t} \\ \hat{\mathbf{v}}_{t+1} &= \frac{\mathbf{v}_{t+1}}{1 - \beta_2^t} \\ \mathbf{W}_{t+1} &= \mathbf{W}_t - \mu \frac{1}{\sqrt{\hat{\mathbf{v}}_{t+1}} + \epsilon} \odot \hat{\mathbf{m}}_{t+1} \end{aligned}$$

図 8: Optimizer Adam

訓練誤差と検証誤差、および正解率を評価指標としてグラフ化し、パラメータやモデル構造の効果を考察する。実装にはPyTorchのnn.Moduleを継承したクラスとしてGraph Layerを定義する。

3.2 結果と考察

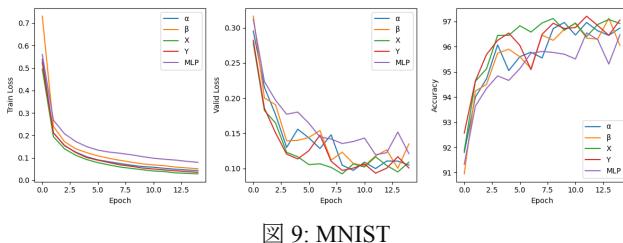


図 9: MNIST

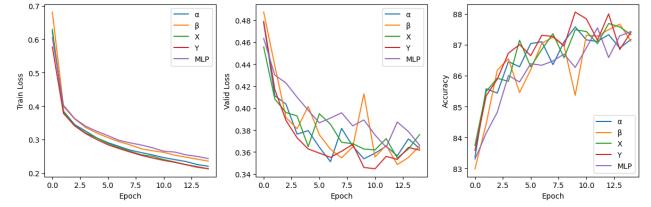


図 10: Fashion-MNIST

MNISTは比較的学習が容易なデータセットであり、全体的に誤差が小さく正解率の高い結果となった。Fashion-MNISTに関しては、MNISTと同様の傾向が見られた。全結合層のみの場では複雑な画像データの分類は難しいが、局所的な特徴量の獲得が可能となった。どちらの結果からもモデルYが最高精度であり、接続距離と再帰回数は実験範囲においては多い方がほしいことがわかった。

モデルを比較し、各パラメータによる影響を考察する。まずは接続距離に関して、 α とX、 β とYを比較すると接続距離は1よりも2の方が精度が高かった。これは、画像を大域的に認識する能力が原因だと考えられるが、この距離を大きくし過ぎることも局所的な値の受け渡しができないため、問題である。次に再帰回数に関して、 α と β 、XとYを比較すると大差はないものの5よりも20の方が精度が高い結果が読み取れる。セグション2.2の図6で言及したが、再帰回数によっては特徴量が失われる可能性があったものの、誤差逆伝播法によるパラメータ更新では情報が失われずに保存されたのだと考える。これは生物の思考でも再帰処理でも特徴の保存が可能だと示唆される。

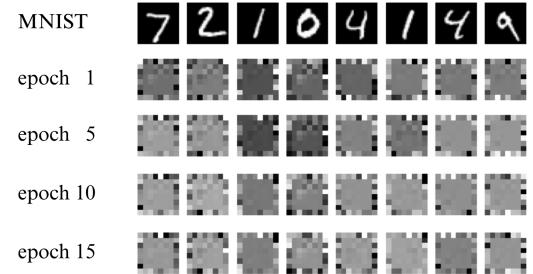


図 11: Graph Layer Output (Model Y)

最終的な検証誤差と正解率が最高値であるモデルYのGraph Layerを可視化した結果を図11に示した。全結合層からの出力が入力されるため、視覚的な特徴を確認することはできないが、同じ数字データに対する出力の近さを調べることは有効である。初めの出力値よりも、最後の方が数字ごとに異なる特徴が見られた。

4. 結論

Graph Layerを用いた実験では特別大きな性能向上は見られなかった。これは、局所的な接続や単純な再帰処理を組み合わせるだけではニューラルネットワークの発展が難しいことを示唆された。しかし、膨大なユニット数における3次元空間での値の送受や再帰回数による制限をなくし、新たに出力を決定する機構を考察することで精度が向上する可能性は残されており、これからも研究されるべき内容だと考える。また、実験を通して最も興味深いことは、再帰回数を増やすと特徴量の消失が懸念されたが、学習ではより高い精度を示した点である。実際のニューロン構造をより鮮明に再現したとしても、既存のニューラルネットワーク構造と最適化手法では高精度化することは難し

いが、機械学習タスクと生物は異なるため、一概に優劣を決定することはできない。数学的にも計算資源的にも非常に困難ではあるものの、根本的な構造改善が必要とされると考えた。ノーフリーランチの定理によれば、あらゆる問題に対して最適なモデルは存在しないとされる。しかしながら、完全に生物の思考プロセスを再現に成功すれば、同様の思考力は得られ、より高度な人工知能が誕生することが期待されるだろう。

結論として、機械学習に対して生物学からの基礎研究アプローチが引き続き行われることが重要であり、そのためには根本的に構造を変化させる必要があると考える。

参考文献

1. Hinton, Geoffrey. (2022). The Forward-Forward Algorithm: Some Preliminary Investigations.
2. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE
3. Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult.
4. Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. International Conference on Machine Learning.
5. Diederik P. Kingma, Jimmy Ba. (2014). Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations.