連載

5分で分かる!? 有名論文ナナメ読み

矢田部浩平（早稲田大学）

Sho Sonoda and Noboru Murata :
Neural Network with Unbounded Activation Functions is Universal Approximator

Applied and Computational Harmonic Analysis, 43(2), pp.233-268 (2017)

ニューラルネットは万能？

深層学習の成功で，ニューラルネットはより一層身近な存在となった．さまざまな応用がされる中で，決めゼリフかのように「ニューラルネットは万能近似能力を有する」などと「万能性」というキーワードが用いられている．万能という単語には「すべてにおいて優れている」という意味もあるので，欠点のない完璧な手段かのように誤解する原因になっている．実際は，理想的には何でも近似可能という，近似器として望ましい性質のことである．

それっぽく言えば「あらゆる関数をいくらでもよく近似できる」のが万能性である．このことは1990年前後にさまざまな理論で示され，今日では常識かのように語られる性質となった．しかし，それらの理論ごとに「どのような種類の関数を何によって」近似できるのか，仮定や結論が異なる．ニューラルネットは「その理論が保証する範囲においては」万能関数近似器だが，それ以外の場合については理解されていないことも多い．たとえば，1989年にCybenkoによって「シグモイドを活性化関数とした」ニューラルネット[☆1]が「超立方体上の連続関数を」近似可能であることが示されているが，この理論は連続でない関数を対象としていない．一般に，性質が良い関数ほど理論を構築しやすいので，万能性に関する理論でも初期のうちは簡単な関数が扱われ，時代とともに徐々に高度化していった．

活性化関数の多様化

活性化関数の選択によってニューラルネットの性質は変わる．深層学習の発展とともにトレンドも変化し，ReLUに代表されるような非有界な活性化関数が当たり前のように用いられるようになった[☆2]．ReLUは無限遠で∞に発散し，また原点における微分が連続でないなど，シグモイドと性質が異なる．当然，それらを用いたニューラルネットも性質が異なるので，異なる理論が必要である．

ReLUのような活性化関数を用いたニューラルネットの万能性はMhaskarとMicchelliによって1992年に示された．その理論では，すでに万能性が知られている他の関数近似手法（具体的にはB-spline[☆3]）を用いてニューラルネットを近似することで，間接的に万能性を示している．この方法はほかの近似手法を介すので，ニューラルネットのパラメータに関する情報が少なく，また有限モデルの構造が保存されるとは限らないという弱点を持っている．ニューラルネットの性質を明らかにするには，そのような間接的な方法ではなく，直接的な理論解析が求められていた．

☆1 ニューラルネットは，活性化関数と呼ばれる非線形関数を用いて定義され，それを足し合わせることで関数を近似する．したがって，活性化関数として何を選ぶかによって，ニューラルネットの関数としての性質が大きく変わる．シグモイド関数は，古典的なニューラルネットでよく用いられる滑らかな活性化関数であり，滑らかな関数の近似に適している．

☆2 ReLU（Rectified Linear Unit）は負の値をゼロで置き換える活性化関数で，深層学習でよく用いられている．入力が大きくなると関数値がいくらでも大きくなれるので，非有界な活性化関数である．

☆3 B-splineは区分的多項式の一種であり，区間ごとに異なる多項式を繋ぎ合わせたものである．関数の近似において多項式はよく用いられるが，区分多項式の方が良い性質を持つ場合が多い．

ニューラルネットの解明に向けて

応用において，ニューラルネットは長らくブラックボックスとして扱われてきた．その構造を明らかにし，ニューラルネットをホワイトボックスにする取り組みの1つが，本稿で紹介する論文である．

タイトルにあるように，この論文は「非有界な活性化関数を用いたニューラルネットの万能性」を示している．具体的には，活性化関数として（多項式関数以外の）緩増加超関数☆4 を選んだ場合に，ニューラルネットが概収束・L^1 収束・L^2 収束の意味で☆5 万能性を持つことが示されている．この理論は，ReLU を含む非常に広い種類の活性化関数を扱うことができ，従来の理論からは見えてこないさまざまな情報を与えることができる．

この論文の特筆すべき点は，間接的な解析手法とは異なり，ニューラルネットを直接的に解析することで「ニューラルネットが何を学習しているか」を明らかにしたことである．実際，緩増加超関数を含む問題設定においても学習パラメータの大域最適解を解析的に与えており，万能近似能力を保証する活性化関数と学習パラメータの組合せも明らかにしている．これは，ニューラルネットの特徴付けや精密評価に繋がる結果である．

積分表現とリッジレット変換

そのような強い結果が得られているのは，ニューラルネットの中間層を無限化して積分作用素として扱い，関数解析や調和解析☆6 的な手法を用いていることによる．通常のニューラルネットは中間層のパラメータが有限個だが，それを無限だと思って連続化し，総和を積分で置き換える．そうやって得られる積分変換は，ニューラルネットの積分表現と呼ばれている．積分で表現することでニューラルネットを線形化でき，また解析学の豊かな知見を利用することもできるので，多くの情報を理論的に引き出すことができる．なお，積分表現は Dirac 測度を用いて直接的に有限モデルも表現でき☆7，現実の有限モデルとの関係性を評価することができる．

積分変換によってニューラルネットを表現する方法自体は，古くから用いられている．この論文の特徴は，Radon 変換とウェーブレット変換を合成した積分変換☆8 である「リッジレット変換」を解析に積極的に用いることで，最適な学習パラメータまでも具体的に計算していることである．リッジレット変換はニューラルネットの積分表現の作用素の右逆作用素なので，学習対象の関数をリッジレット変換すれば，最適な学習パラメータを具体的に与えることができる☆9．この論文では，非有界な活性化関数を用いた場合でもリッジレット変換が構成でき，再構成公式が成り立つことを示している．これにより，ReLU のような活性化関数も積分表現とリッジレット変換の枠組みで扱うことができるようになり，強力な理論解析の恩恵を受けられるようになった．

そして博士論文がバズる

本稿で紹介した論文は，中間層が1層の浅いニューラルネットを対象としている．積分表現その

☆4 関数の一般化として扱われる「超関数」の中で，比較的性質の良いもの．たとえば Dirac の δ 関数などが有名だが，二乗可積分関数や多項式関数など通常の関数も含む．緩増加超関数は増大度が高々多項式程度でなくてはならないが，活性化関数として用いられる関数のほぼすべてを含むと考えて問題ない．

☆5 近似対象のあらゆる関数に対して近似誤差をゼロにできることが万能性だが，近似誤差をどのように定義するかによって意味が変わってくる．この論文では，3種類の定義に対して万能性を示している．

☆6 調和解析は Fourier 解析などを含む数学の一分野であり，積分変換によって関数を表現する方法などについての理論が研究されている．

☆7 積分は「足し合わせること」を理想化したものと見なせるが，Lebesgue 積分では「どのように足し合わせるか」を測度の選び方によって定めることができる．Dirac の δ 関数のようなもので都合よく定義すれば，総和をそのまま積分で表現できる．

☆8 X 線 CT などで用いられる Radon 変換や，画像処理などで用いられるウェーブレット変換といった有名な変換と比べて知名度が低いので，リッジレット変換はそれらの合成として紹介されることが多い．

☆9 関数をリッジレット変換し，その変換結果（リッジレットスペクトル）を重みとしてニューラルネットの出力を計算すると，元の関数が再構成される．しかし，ニューラルネットを表す積分変換は全射だが単射でないので，リッジレット変換は逆作用素ではなく右逆作用素である．つまり，逆の順番で，先にニューラルネットの出力を計算してからリッジレット変換をしても，元のスペクトルは得られない．

ものでで深層ニューラルネットを直接扱うことはできないので，深層学習に応用するには追加の工夫が必要となる．この論文の著者は，微分方程式や力学系に基づいて深層ニューラルネットを定式化することで，積分表現の深層学習への応用も行っている．

これらの内容は，博士論文として日本語でも出版されている[1]．この博士論文こそが「有名論文」であり，SNS でバズったことで日本における積分表現理論の知名度を飛躍的に上げた論文である．実際，本稿執筆時点で 77,394 回ダウンロードされており，博士論文のダウンロード数としてはトップクラスの多さだと思われる．導入にはニューラルネットに関する話題が幅広く含まれており，当時のサーベイ論文としても興味深い．無料で PDF がダウンロードできるので，気軽に覗いてみるとよいかもしれない．

参考文献

1）園田　翔：深層ニューラルネットの積分表現理論，早稲田大学博士論文（2017）．

（2021 年 1 月 29 日受付）

矢田部浩平
k.yatabe@asagi.waseda.jp
2012 年学部卒業，2014 年修士修了．光を用いて空中の音を見る研究により 2017 年に博士号を取得．2017 年より早稲田大学表現工学科助教，2018 年より同学科講師（任期付）．MATLAB を好んで用いる．