



Geoffrey E. Hinton, Simon Osindero and Yee-Whye Teh : A Fast Learning Algorithm for Deep Belief Nets

Neural Computation, Vol.18 Issue 7, pp.1527-1554 (July 2006)

深層学習のもとになった?

深層学習 (deep learning) への大きな注目は言を俟たないが、そのきっかけはこの論文であると考えられる人は多い。Ian Goodfellow らの (恐らく世界で最も使用されている) 深層学習の教科書¹⁾ や、学術的な解説記事にもそのような記述がある。しかし、この論文の原著を読んだ人は少ないかもしれない。なぜなら、深層学習が爆発的に広まったのは、Krizhevsky, Sutskever, Hinton の 2012 年の画像分類コンペでの優勝²⁾ あたりからであり、今ではこの論文 (とこれに続く事前学習のシリーズ) のテクニックはあまり利用されることがないからである。そこで、この技術的なギャップを紐解くべく、この論文の内容、貢献、およびその後の展開についてお伝えしてみたい。

ざっくりと言ってしまうと、この論文で Hinton らは、深層信念ネットワーク (deep belief nets) で層を深く (具体的には、隠れ層の数を三層に) することに成功した。それまでは隠れ層 (中間の層) の数はたかだか一層か二層であり、層の数を増やすと汎化性能の高い解に辿りつかせることが難しかった。詳細は後述するが、学習が成功したポイントは、ネットワークを層ごとに教師なしで事前学習させておく、というものである。これを層の数だけ行っておけば、後の全体的な教師あり (または教師なし) 学習で良い結果が得られることが示されている。

深層信念ネットワークとその事前学習

深層信念ネットワークは、ニューラルネットワークと類似した、図-1 左に示すようなネットワークである。図-1 左の白丸がユニット、白丸の横の並びが層、黒線が結合であり、結合には重みが付与されていて、つながりの強さを表す。ニューラルネットワークと少し異なるのは、ユニットが実数でなく 0 または 1 の値を確率的にとる点である。また隠れ層の最上層が連想記憶 (associative memory) と呼ばれ、青丸で示された最下層の入力層へ推論がつながる生成的なモデルとなっている。ユニットが 1 をとる確率は、前層の出力に結合重みをかけて、総和をとったものからバイアスを差し引き、これをシグモイド関数に入力して決められる。深層信念ネットワークでは連想記憶との結合が無向で、それ以外は上から下へ流れる有向な結合となる。

論文では、図-1 右に示すようなネットワークを 10 種類の手書き数字 (0 から 9 まで) の分類に適用した。28 × 28 ピクセルの画像を白と黒に二値化し、これを 784 個のユニットにしたものが最下層の入力層である。隠れ層は三層で構成されており、全体のパラメータ数は 166 万個を超える。

ネットワークが所望の出力をするように、学習データを使ってパラメータを最適化したいのであるが、これほどパラメータ数の多いモデルでは、ランダムな初期値からの学習は難しかった。そこで論文

で提案されているのは、下層から一層ずつ取り出し、制約付きボルツマンマシン^{☆1}とみなして、下層からの入力を再生成するようにパラメータを事前学習することである。画像が何であるかを示すラベルの情報が必要ないので、このような学習を教師なし学習と呼ぶ。一層目が学習できたら、その出力を次への入力とみなして、次層の学習を行い、これを最上層に到達するまで繰り返す。その後、微調整 (fine tuning) といって、教師信号 (たとえば数字が何であるかのラベル) を用いて全体を最適化する^{☆2}。

結果として、6万枚の訓練画像で学習し、1万枚のテスト画像で数字を当てさせたときのエラー率が1.25%であった (つまり125枚間違えた)。ニューラルネットワークで隠れ層が二層のものや、当時流行していたカーネル法を用いたサポートベクタマシン

よりも若干エラー率が向上した。しかし、1998年のLeCunらの畳込みニューラルネット (LeNet-5) が使用したパラメータの数は6万個で、エラー率は0.95%であったことを考えると、この結果そのものよりも、このような規模のネットワークで学習が成功したことにインパクトがあった。

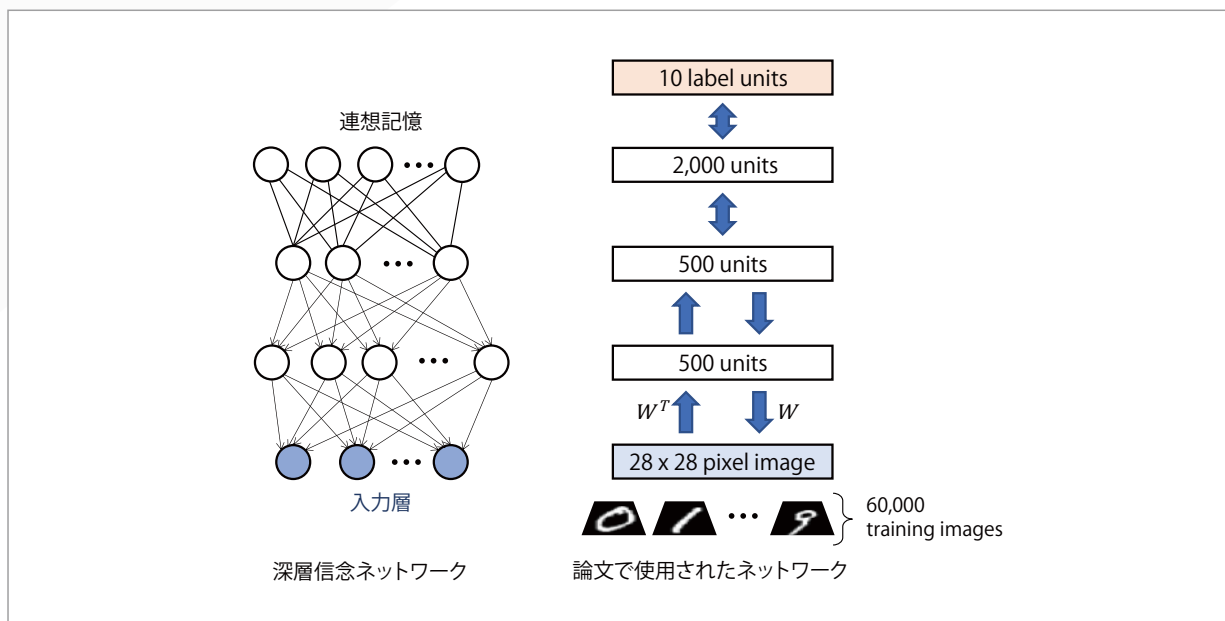
その後の経緯とこの論文の位置づけ

Hintonらは同時期に、自己符号化器 (auto encoder) でも同様の結果を得ている。その後もいくつか、類似した事前学習やなぜそれが上手くいくかを考察した論文が発表され、少なくともGoogleの猫^{☆3}くらいまでは、自己符号化器では層ごとに教師なし学習をするのが前提になっていたようである。しかし、このような事前学習が有効であることは理論的な示唆はあっても、実験的にしか示されていなかった。

☆1 このとき、上から下への結合重みを W とすれば、逆方向の重みは W の転置が使われる。制約付きボルツマンマシンや事前学習のテクニックに関しては、深層学習の教科書^{1), 3)} や岡谷らの解説記事⁴⁾ が詳しい。

☆2 制約付きボルツマンマシンの学習には、ギブスサンプリングよりも収束がはやいコントラスト・ダイバージェンス法が用いられた。また微調整には、上層へのパスと下層へのパスを繰り返しながら、それぞれで誤差が小さくなるように各方向のパラメータを更新するWake-Sleep アルゴリズムに似た最適化が用いられた。

☆3 Google と Stanford のチームが自己符号化器を積層してYouTubeで収集したビデオのフレームから教師なし学習を行い、猫や人の顔に反応する局所結合の重みの学習に成功した。このとき、3つの積層された自己符号化器において、パラメータの最適化は同時に行われたが、それぞれが下層からの入力を再生成させる誤差関数が設定されていた。



◆ 図-1 深層信念ネットワーク

その後、現在までの間に、(ご承知の通り)、いくつかの重要な(再)発見^{☆4}が分野の進展を加速させた。その結果、現在では、確率的なモデルよりも扱いが容易な確定的なネットワークが主流になった。また、自己符号化器や生成的なネットワークを含む階層の深いネットワークはさまざまに利用されるものの、数百万~数千万個のパラメータがあっても、一部だけを取り出して最適化するような事前学習はされないのが一般的である。そこで改めてこの2006年の論文を顧みるに、ネットワークを深くするという目標の下、先駆けとなった研究者らが試行錯誤していた中で、上手くいった第一歩と捉えるのがよいと思う。真に技術的な意味でのブレイクスルーは AlexNet²⁾ だと言う人や ResNet⁵⁾ (あるいはさらに別の論文) だと言う人もいるかもしれない^{☆5}。いずれにしても、振舞いが確率的であるために実装上は皆敬遠しがちであるが、理論的には解析しやす

い深層信念ネットワークのようなモデルで最初の一歩があったことは興味深い。また、誤差逆伝播法、深層信念ネットワーク、制約付きボルツマンマシン、コントラストティブ・ダイバージェンス法、Wake-Sleep アルゴリズム、ReLU, AlexNet, これらすべてが Hinton と共著者の著作であり、彼らの一貫した研究が2012年の画像分類コンペでの優勝につながったことには敬服せざるを得ない。

参考文献

- 1) Goodfellow, I., Bengio, Y. and Courville, A. : Deep Learning, MIT Press (2016).
- 2) Krizhevsky, A., Sutskever, I. and Hinton, G. E. : ImageNet Classification with Deep Convolutional Neural Networks, NIPS (Dec. 2012).
- 3) 神島敬弘 編, 人工知能学会 監修: 深層学習, 近代科学社 (2015).
- 4) 岡谷貴之, 齋藤真樹: ディープラーニング, 情報処理学会研究報告 (2013).
- 5) He, K., Zhang, X., Ren, S. and Sun, J. : Deep Residual Learning for Image Recognition, CVPR (June 2016).
(2018年7月20日受付)

☆4 たとえば、GPUによる高速な並列計算、効率的な学習率の設定法、畳込み、プーリング、ReLU (Rectified Linear Unit) によるシグモイドの代替、結合を確率的に切る Dropout などの正則化、学習データの拡張、コントラスト正規化、層を飛ばすスキップコネクション、などである。

☆5 現時点で、本論文が8千回、AlexNet²⁾ が2万6千回、ResNet⁵⁾ が1万回の引用回数である。これらがいかに多いかといえば、Thomson Reuters の調査でDNAのらせん構造の発見が5千回強、オゾンホールの発見が2千回弱の引用回数である。

.....
川上 玲 (正会員) rei@hc.ic.i.u-tokyo.ac.jp

2008年東京大学大学院博士課程修了。博士(情報理工学)。同大学院情報学環などを経て、2018年より同大学院特任講師。コンピュータビジョンに関する研究に従事。

