Ⅰ AIを知る―― AIとは何か，何を変えるのか

4. ディープラーニングの技術開発の現状と展望

井﨑 武士 エヌビディア合同会社エンタープライズ事業部

2012年，米国Google社がネコを認識するAIを発表して以来，ディープラーニングの研究および応用は急速に進んでおり，インターネットやクラウドの分野のみならず，メディア，金融，セキュリティ，製造業など多くの分野で活用事例が報告されている。この数年でディープラーニングが加速してきた要因は，3つあると言われている（図1）。

1つは，アルゴリズムの進化である。かつて，1980年代には，誤差逆伝播学習法を用いた多層パーセプトロンの研究が盛んに行われていた。しかし，勾配消失問題や過学習を防ぐ手法が確立されておらず，実用化に至るまでの精度を出すことができなかった。2010年頃の活性化関数ReLUの導入や，ドロップアウトにより実用化に耐えうる精度が出せるようになった。2つ目は，ビッグデータである。ニューラルネットワークの学習で精度を上げるためには，大量の学習データが必要となる。現在インターネットなどを通じて，ビッグデータと呼ばれる大量の学習データが入手可能となった。そして，3つ目はGPUである。大量のデータを用いて学習を行う際，CPUを使用していては開発期間が長期化し，実用化が困難であった。GPUを用いて並列処理を行うことで，開発期間が現実味のある期間となった。

2015年のImageNet Large Scale Visual Recognition Competition（ILSVRC）で1位となったMicrosoft Research Asia（MSRA）が使用したニュートラルネットワークのモデルである“ResNet”[1]は，約6000万パラメータを有し，約700京回の計算量を必要とする。最新のCPUでも，3か月程度は要する。しかし，最新のGPUを用いれば，1週間程度で演算可能である。ただ，実際には，1回の試行でパラメータの調整は完了せず，ネットワークの構成などを変えながら何度も試行されるため，計算能力に対する要求は非常に高い。通常は，1ノードに複数のGPUを搭載し，また，複数ノードに分散して試行期間をさらに短縮化している。

ただ，年々モデルは複雑さを増しており，パラメータや演算量は飛躍的に増大している。そのため，さらなるハードウエアの高性能化が求められている（図2）。

エヌビディア ディープラーニング プラットフォーム

ディープラーニングには，2つのフェーズがある。モデルを作りこむ「学習」，およびそのモデルを使用する「推論」である。学習には，ワークステーションやサーバが用いられ，エヌビディアにはその用途向けのGPU製品，「Quadro」や「Tesla」がある。推論の場合，オンライン環境で行うのであれば，そのままサーバなどを使用可能だが，オフライン環境であれば，システムオンチップ（以下，SoC）（エヌビディアのGPU搭載のSoC組み込みモジュール「Jetson」および車載用の「DRIVE PX 2」）が利用可能である。デバイス内のGPUアーキテクチャは同一のため，同じCUDAコードが実行可能である。エヌビディアでは，約2年ごとにアーキテクチャの更新を行っており，毎度消費電力効率を約2倍に向上してきている。現在入手可能な製品の最新アーキテクチャの

図1 ディープラーニングを加速した3要因

〈0913-8919/17/￥300/ 論文 /JCOPY〉