



# インテルが推進するディープ ラーニング・フレームワーク

さらなる洞察を提供

Pubudu Silva インテル コーポレーション マシンラーニング / コンピューター・ビジョン シニア・ソフトウェア・エンジニア

人工知能 (AI) は、本来は人の知能が必要な、視覚的理解、音声認識、言語処理、意思決定のようなタスクを実行することができる知能機械 (コンピューター) の概念であり、コンピューターの登場以来、次の大きな挑戦であると考えられています。

マシンラーニングは、いくつかの重要な AI タスクを実行するのに非常に有効であることが分かっています。人工ニューラル・ネットワーク (ANN) は、哺乳類の脳皮質の神経細胞構造の数学モデルであり、その遠大な設計とさまざまなタスクに対する汎用性により、特に AI 向きであると考えられていました。ANN の強みは、隠れた過渡状態 (隠れたノード) を学習して保守する能力にあります。この能力により、いくつかの非線形関数をカスケードして、入力から目的の出力まで、広い範囲のマップを学習することが可能です。

学習済みの ANN では、隠れた層は階層的な段階でデータの内部的な抽象化を表します。深い層ほど、より高いレベルの抽象化を表します。哺乳類の脳は複数の階層的な処理層で情報を処理すると考えられています (例えば、霊長類の視覚システムでは、処理が連続した段階で行われ、エッジ検出からプリミティブ形状検出を経て、徐々に複雑な視覚形状に移動します)<sup>1</sup>。したがって、AI 研究には多層の「深い」ANN が望まれます。

深いカスケードされた層を備え複数の連続した段階でデータを処理するネットワークは、一般に「ディープ・ネットワーク」と呼ばれます。サポート・ベクトル・マシン (SVM)、混合正規分布 (MoG)、k 近傍法 (kNN)、主成分分析 (PCA)、カーネル密度推定 (KDE) のような、最も広く利用されているマシンラーニング・アルゴリズムには、3 つを超える処理層は含まれていません。したがって、これらは「浅い」アーキテクチャーと見なすことができます。2 層から 3 層の ANN は、うまくトレーニングすることができます。20 世紀の終わりに、より深い ANN をトレーニングするいくつかの試みが行われましたが、次の 2 つの問題に直面したことで失敗に終わりました。

1. 勾配消失問題
2. 多層化による重みの増加から生じる過学習

コンピューティングの進歩とともに、研究者は短時間に膨大なデータサンプルを使用してマシンラーニング・アルゴリズムのトレーニングを行うことが可能になり、過学習問題はほぼ解決しました。畳み込みニューラル・ネットワーク (CNN) は、多層型のディープ・ネットワークですが、重み共有の概念により (同等の深さの全結合 ANN よりも) 小さな多くの重みを含んでいます。このため、CNN は ANN よりもはるかに容易にトレーニングを行うことができます。CNN の理論的なパフォーマンスは ANN にはわずかに及びませんが、教師あり画像認識タスクでは最も普及している手法です。2006 年のジェフリー・ヒントン教授などによる画期的な提案<sup>2</sup> のおかげで、ディープ・ビリーフ・ネットワーク (DBN) やディープ・オート・エンコーダーのようなディープ・ネットワークの手法は教師なし学習にも利用されています。

一般に、ディープ・ネットワークは、分類、回帰、画像認識、自然言語処理などのほとんどの AI 関連タスクで、ほかのマシンラーニング・アルゴリズムよりも高いパフォーマンスを示します。ディープ・ネットワークの大いなる成功は、特徴階層の自律学習によるものです。

ディープラーニングと従来の統計的学習手法の大きな違いは、前者はローデータを学習するのに対して、後者はデータの間工学的特徴を学習する点です。DNN 学習の初期レベルで、ディープ・ネットワークは、指定されたタスクに適した特徴を自律的に生成します。目的関数に基づく最適化プロセスが、指定されたオリジナルデータを用いてすべての学習タスクを行うため、学習プロセスから推測や人間による判断の偏りが取り除かれます。

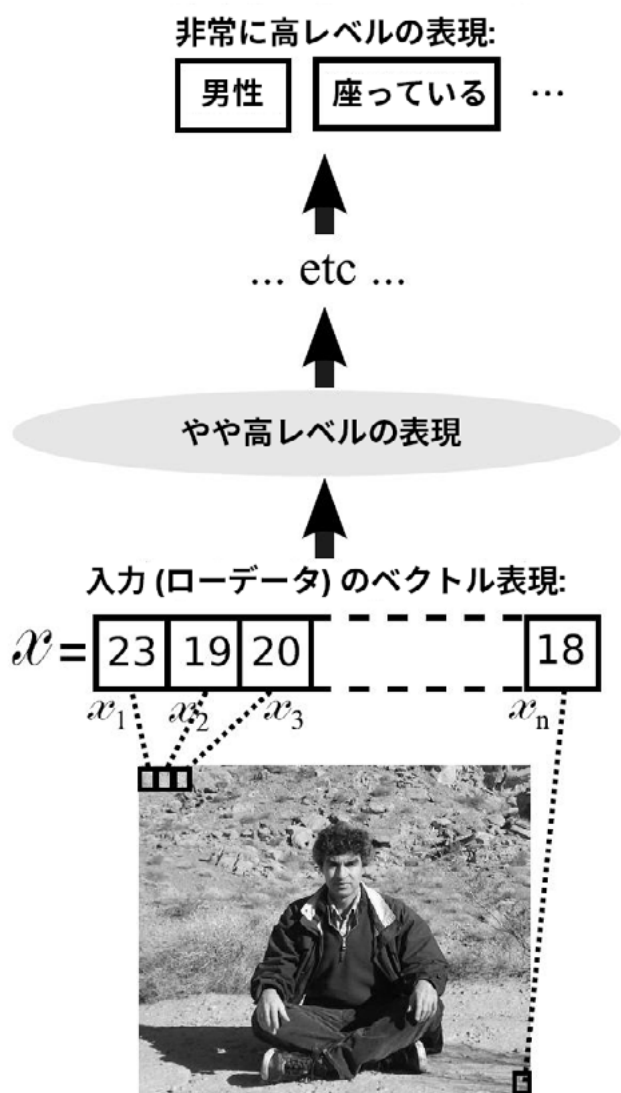
深い階層構造により、特徴の階層的な学習 (ネットワークの前のレベルで学習した低いレベルの特徴に基づいて、深いレベルではより高いレベルの特徴を学習する) が可能になります。

図 1 は、入力画像が徐々にディープ・ネットワークのより高いレベルの表現に変換される様子を示しています。ネットワークの深くに移動するごとに、画像はより抽象的な表現になります。例えば、特徴の階層的な学習は、初期の層から深い層へ、エッジ、形状、オブジェクトの一部、完全なオブジェクト、シーンの順で行われます。しかし、実際には、すべてを学習することなく、これらの階層的な抽象化層の「正しい」特徴ベクトルを推測することは困難です。これは、人間工学的特徴を学習する際に問題となります。一般に、深いネットワークでは、

出力層は非常に高レベルの特徴を処理して、浅いネットワークよりもはるかに高いレベルの概念を学習できるようにします。

さまざまなディープラーニングの手法が開発者、データ・サイエンティスト、研究者の標準ツールになったように、ディープ・ネットワークのトレーニングとスコア付けを支援するため、多くのディープラーニング・フレームワーク (Caffe\*、Tensorflow\*、Theano\*、Torch など) およびライブラリー (MatConvNet、CNTK、Pylearn2、Deeplearning4j) が開発されています。これらのフレームワークとライブラリーは、退屈で単純な作業を減らすのに大いに役立ちます。ユーザーは、個々のコンポーネントの実装ではなく、ディープラーニングのさまざまな面に労力をかけることができます。

さらに、フレームワークとライブラリーは開発者コミュニティの協力の下で、オープンソース・プロジェクトとして開始されることが多いため、ユーザーはフレームワークとライブラリーのコードベースにアクセスすることができます。ディープラーニングでは、完了まで数日から数週間かかる、非常に大きなデータセットでトレーニングが行われるため、一般的に使用されるディープラーニング・ソフトウェアのパフォーマンス最適化が非常に重要です。



1 入力画像を徐々に変換してより高いレベルの表現にするディープ・ネットワーク処理  
 Source: Yoshua Bengio, Learning Deep Architectures for AI, 2009

インテルは、特にインテル® アーキテクチャー向けに最適化することにより、さまざまなオープンソースのディープラーニング・フレームワークの普及に貢献しています。[マシンラーニング・サイト](#) (英語) では、インテルによるマシンラーニングとディープラーニングの取り組みに関する最新の情報を提供しています。パフォーマンス最適化ツールと手法の詳細も、このサイトで提供しています。最適化に関するいくつかの情報は、独自のディープラーニング・アプリケーションを開発する際、およびフレームワークやライブラリーを開発サイクルで使用する際にソフトウェア開発者の役に立つように、ケーススタディーとして公開されています。

例えば、Caffe\* の最適化で行われたプロセスは、[インテル® アーキテクチャー向けに最適化された Caffe\\*: モダンコード手法の適用](#) (英語) で示されています。[インテル® VTune™ Amplifier XE](#) は、CPU、キャッシュ、CPU コア、メモリーの使用状況、スレッドのロードバランス、スレッドのロックなど、パフォーマンス最適化の初期ガイダンスとして使用できる、さまざまな洞察を提供する強力なプロファイルツールです。[インテル® マス・カーネル・ライブラリー \(インテル® MKL\)](#)、[インテル® スレディング・ビルディング・ブロック \(インテル® TBB\)](#)、OpenMP\* のようなライブラリーは、ディープラーニング・ソフトウェアの最適化に非常に役立つことが分かっています。

ディープラーニングの開発と研究を促進するため、インテルは、ディープラーニング・ソリューションの開発、トレーニング、配置を支援する、データ・サイエンティストとソフトウェア開発者向けの無料のツールセットとして、[インテル® Deep Learning SDK](#) (英語) を提供しています。インテル® Deep Learning SDK は、Ubuntu\*/CentOS\* サーバーに接続された Web ベースのクライアントとして設計されています。インストール・ウィザードは、SDK とサーバーのインテル® アーキテクチャー向けに最適化された一般的なディープラーニング・フレームワークをインストールします。トレーニング・ツールは、単純なグラフィカル・ユーザー・インターフェイスと高度な視覚化手法により、トレーニング・データの準備、モデルの設計、モデルの選択を大幅に単純化します。配置ツールは、モデル圧縮と重み量子化手法により、特定のターゲットデバイス向けにトレーニングされたディープラーニング・モデルを最適化します。

## 参考資料 (英語)

1. Serre, T.; Kreiman, G.; Kouh, M.; Cadieu, C.; Knoblich, U.; and Poggio, T. 2007. “A quantitative theory of immediate visual recognition.” *Progress in Brain Research, Computational Neuroscience: Theoretical Insights into Brain Function*, 165, 33–56.
2. Hinton, G. E.; Osindero, S.; and Teh, Y. 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554.

## さらなる洞察を提供

[intel.com](#) (英語) では、インテル向けに最適化されたディープラーニング・フレームワーク、ライブラリー、インテル® Deep Learning SDK のようなディープラーニング・ツールについての最新情報を提供しています。

**インテル® DEEP LEARNING SDK と  
その他の AI ソフトウェア・ツールの詳細 (英語) >**

# インテルの技術ウェビナー (英語) で 最新の情報を入手

インテル® ソフトウェアのエキスパートが提供する最新のヒントとコツでコードのパフォーマンスを向上しましょう。ホットなトピックについて取り上げた無料の技術ウェビナーをぜひご覧ください。

インテル® Parallel Studio XE 2017:  
高速なコードを素早く開発、

対応ハードウェアの有無に関係なく  
インテル® AVX-512 向けに最適化、

アプリケーションを共有 / 分散  
メモリーでスケールアップ、

マシンラーニングにおける Python\*  
パフォーマンスの壁を乗り越える、

インテル® Xeon Phi™ プロセッサ  
の高帯域メモリー対応のコード、

ベクトル化: 活用すべき "別の" 並列  
処理、

ルーフライン解析: パフォーマンス  
最適化のトレードオフを視覚化する  
新しい方法、

最新のインテル® プロセッサに  
対応したメディア・ソリューション /  
アプリケーションの開発、

インテル® ライブラリーによる  
ディープラーニングとマシンラー  
ニングの促進、

インテル® HPC オーケストレーター:  
インテル® スケーラブル・システム・  
フレームワークのビルディング・ブ  
ロックを提供するシステム・ソフト  
ウェア・スタック、

2017 年春に公開予定の新しい技術ウェビナーにご期待ください。  
過去の技術ウェビナーは[こちら \(英語\)](#) でご覧になれます。

コンパイラーの最適化に関する詳細は、[最適化に関する注意事項](#)を参照してください。

© 2017 Intel Corporation. 無断での引用、転載を禁じます。Intel、インテル、Intel ロゴ、Xeon、Intel Xeon Phi は、アメリカ合衆国および / またはその他の国における Intel Corporation の商標です。