

Deep Learning概説 –AIの核となる機械学習技術の最先端–

FUJITSU JOURNAL / 2018年11月29日



最近、新聞やテレビでよく目にするDeep Learning。「名前は聞いたことがあるが、具体的に何ができて、どう使われているのだろう」「調べてみたが、数式が多くてよくわからない」そんな人が多いのではないのでしょうか。本コラムではDeep Learningの基本的な仕組みや応用について具体例を交えながら解説します。

執筆者プロフィール

佐藤 文孝（さとう ふみたか）

株式会社富士通総研 ビジネスアナリティクスグループ シニアコンサルタント



データサイエンティストの視点により、様々な課題をデータ分析で解決に導くAI活用のコンサルタント。

執筆者プロフィール

堀田 一真（ほりた かずま）

株式会社富士通総研 ビジネスアナリティクスグループ



製造業を中心に、自然言語や画像、動画などを対象とした組織内データ活用のモデリング・システム開発に従事。

※この記事は、富士通総研発行の情報誌「知創の杜 2018 Vol.2」（2018年3月19日発行）に掲載されたものです。

※執筆者の部署、役職と記載内容は、「知創の杜」発行当時のものです。

Deep Learningとは

そもそもDeep Learningとは何かを簡潔に定義しておきましょう。Deep Learningとは機械学習の手法（アルゴリズム）の1つであるニューラルネットワークを多層構造に構築したものです。機械学習とは人間が行っている学習機能をコンピュータで実現しようという技術であり、昨今のAIと名のついているシステムではこの機能が核になっている場合が多いです（必ずしも機械学習が用いられているわけではありません）。機械学習によりコンピュータはデータの中に潜む特徴を学習し、あたかも自ら成長しているように振る舞ったり、新たな未知のインプットに対応したり、時

には人間の感覚を超えたアウトプットを見せます。機械学習自体は古くからある技術であります
 が、Deep Learningはこの機械学習の限界値を大きく引き上げました。よく混同されやすいAI、機
 械学習、Deep Learning等の関係を図1に示します。

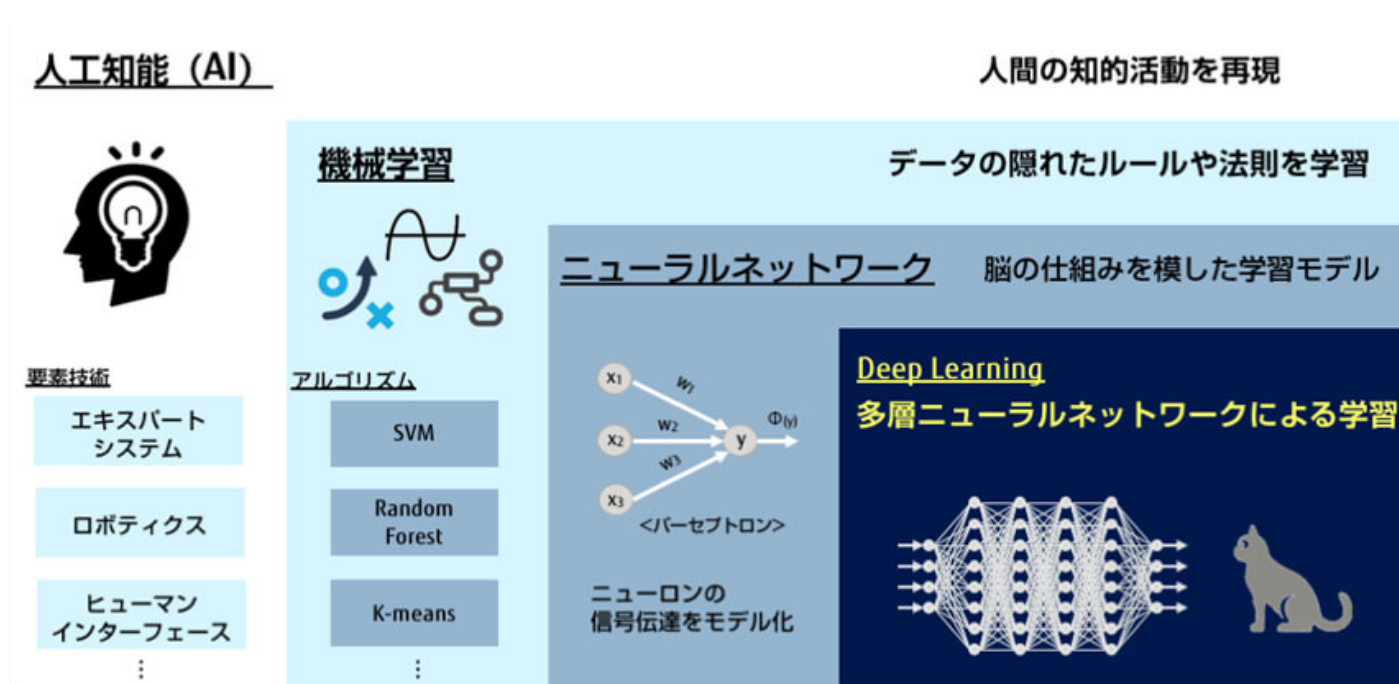


図1 AI、機械学習、ニューラルネットワーク、Deep Learningの関係

機械学習：人工知能の要素技術の1つ

ニューラルネットワーク：機械学習のアルゴリズムの1つ

Deep Learning：多層に構築されたニューラルネットワーク

どのようなアルゴリズムか

突然ですが、以下の画像を見たとき、あなたは何を考えるでしょう？「白と黒の縞模様」「たてがみ」「大きな耳」などの特徴から無意識に「シマウマ」を認識するのではないのでしょうか（図2）。



図2 認識対象とする画像の例（シマウマ）

「シマウマ」の画像。多くの人間は画像を見たときに、画像が「シマウマ」を描画していると無意識に認識するが、実際にはニューロンと呼ばれる神経細胞のネットワークにおいて認識プロセスが行われている。

人間が無意識のうちに「シマウマ」を認識するまでには、実際には以下のプロセスをたどっています。

1. これまでに図鑑やテレビ、動物園等で「シマウマ」を目撃。
2. 脳の中で「白と黒の縞模様」、「たてがみ」、「大きな耳」などの特徴を抽出し、対象が「シマウマ」であると記憶。
3. 図2を見たとき、画像の特徴とこれまでに記憶した様々なものの特徴を照らし合わせ、最も類似するものを関連付け、対象が「シマウマ」であると判断。

実際には、人間は脳内でこれらのプロセスをニューロンと呼ばれる神経細胞のネットワークを用いて行っています。人間の脳内にはニューロンが10¹⁰（10の10乗）個以上あり、各ニューロンがシナプスと呼ばれる接合部位によってつながっています。ニューロンは入力される電気信号の閾値がある一定の量を超えると発火し、シナプスによって次のニューロンに電気信号を出力します。脳はこの動作の連続により信号の伝達を行い、「シマウマ」の認識を含めた様々な思考のプロセスを実現しています。そして脳内の神経回路網とそのプロセスを模倣してコンピュータ上で再現したのが、ニューラルネットワークという機械学習ロジックです。

先ほどの「シマウマ」の認識を例に見てみましょう。ここでは、対象がシマウマかどうか（シマウマの場合は1、そうでない場合には0）の判断に「白と黒の縞模様： x_1 」「たてがみ： x_2 」「大きな耳： x_3 」という3つの特徴を入力として用います。各入力（各特徴のもっともらしさを0~1の値で表現）はそれぞれの比重（シマウマを認識するうえでの各特徴の重要度を0~1の値で表現）を掛けたうえで足し合わされ、その合計が閾いき値を超えていれば出力は1、そうでなければ出力は0となります。例えば、各特徴の入力を $x_1=0.9$, $x_2=0.6$, $x_3=0.8$ 、重みを $w_1=0.8$, $w_2=0.5$, $w_3=0.6$ とすると、 $0.9 \times 0.8 + 0.6 \times 0.5 + 0.8 \times 0.6 = 1.5$ を計算して、閾値 θ と比較する。仮に閾値を $\theta=1.0$ とすると 1.5 （計算結果） > 1.0 （閾値）となり、出力は1（画像が「シマウマ」を描画している）となります（図3）。今回は各入力の重要度を表す重みの値を $w_1=0.8$, $w_2=0.5$, $w_3=0.6$ と定義しましたが、実際にニューラルネットワークを利用する際には、シマウマを認識するうえで最適な重みを大量のシマウマの画像から統計的に探索していきます。このニューラルネットの重みを入力データから最適化していくプロセスのことを学習と呼びます。

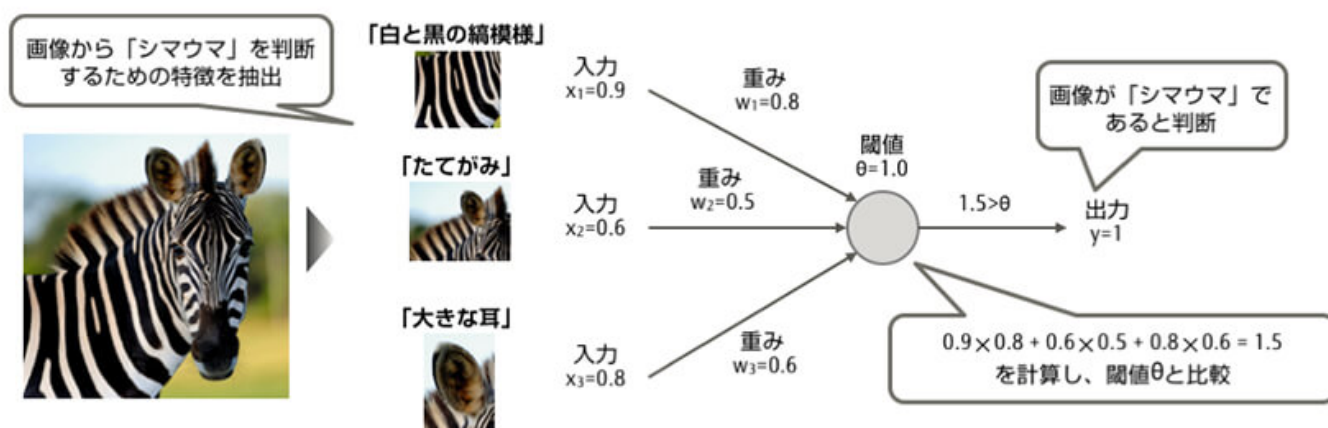


図3 ニューラルネットワークによる画像認識ロジック

機械学習の手法の一つであるニューラルネットワークにより、「シマウマ」を認識するまでの過程。対象がシマウマかどうかという判断に「白と黒の縞模様」、「たてがみ」、「大きな耳」という3つの特徴を入力とし、それぞれの比重を掛けたうえで足し合わせる。合計が閾値を超えていれば対象の画像が「シマウマ」と判断する。

このように、脳内の神経回路網とそのプロセスを模倣し、コンピュータ上で再現したものがニューラルネットワークという機械学習ロジックであり、これを多層にして組み合わせたものの総称を「Deep Learning」と呼びます。シマウマの例では、各特徴とその入力の値を例として定義しましたが、「何についての特徴」を入力とし、それが「どのくらいもっともらしいのか」という点についてもニューラルネットワークで学習してしまおう、というのがDeep Learningの発想であり、こここそが、いかに画像から有用な特徴量を抽出するかについて人手で設計する従来の画像認識手法と大きく異なる点です。このプロセスを繰り返すことによって、最終的には画素単位の情報をもとに画像がシマウマかどうかという判断が可能となります。また、ニューラルネットワークの出力を図4の出力層のように複数設けることで、対象がシマウマなのか、猫なのか、といった、複数の動物の中から対象を識別する複雑な問題についても対応が可能となります。

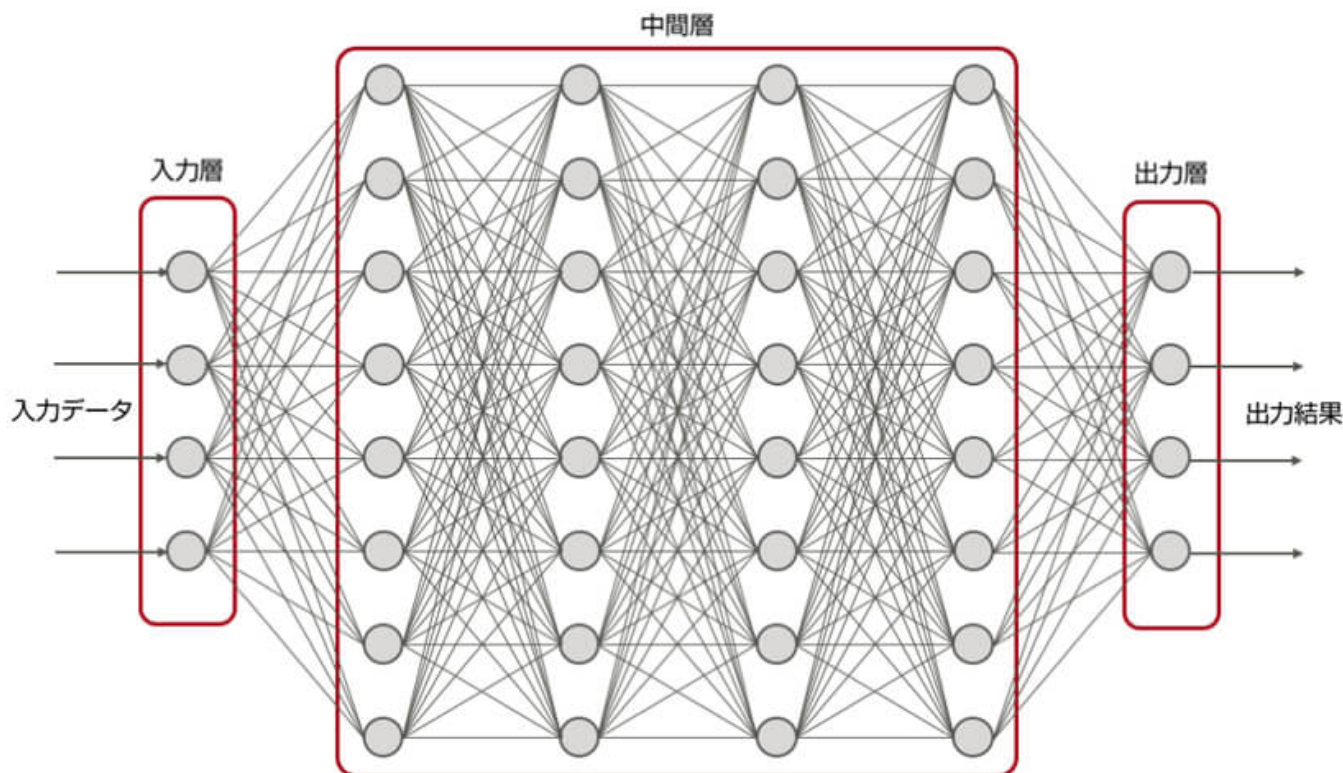


図4 Deep Learningの構成イメージ

中間層にニューラルネットワークを積み重ねることにより、画素単位の情報をもとにシマウマを認識することが可能となる。また、ニューラルネットワークの出力を上記のように複数設けることで、対象がシマウマなのか、猫なのか、といった、複数の動物の中から対象を判別する複雑な問題についても対応が可能となる。

AI囲碁、画像認識、言語処理など、様々な分野で発展するAI

Deep Learning技術の発展はそれまでの機械学習の限界値を大きく引き上げ、第三次人工知能ブームの原動力となったことは確かです。今までのコンピュータの主な役割は決められた処理を高速にこなせることでありました。現在のAIとしてのコンピュータは、ある部分においては人間の判断を代替したり、時には人間が気づかない事象を発見したりすることを期待されるようになってきています。

前出したDeepMind社が生み出したAI囲碁ソフト「アルファ碁」がトッププロ棋士に圧勝したニュースは人々の記憶にまだ生々しく残っているでしょう。もはやルールが決められた知的ゲームでは人間は機械には勝てなくなってしまうと断言していいでしょう。このアルファ碁もプロ棋士同士による棋譜16万局、3,000万盤面の学習にDeep Learningを用いています。

画像認識でいえば、2012年画像認識の精度を競うコンテストであるILSVRC (ImageNet Large Scale Visual Recognition Challenge) において、トロント大学のHintonグループは8層の畳み込みニューラルネットワークを用い、1,000クラス識別の誤り率で2位以下のチームに10%以上もの大差をつけて圧勝しました。それ以降、画像認識の手法は、人による特徴量設計に主眼を置いた従来のアプ

ローチからニューラルネットワークを用いた手法へと一気に移行し、2015年にはついに人間の認識率を上回る結果を出すことに成功しました。

言語処理においては2016年にGoogle翻訳がDeep Learningを用いた翻訳モデルにアップデートされ、劇的な精度向上を果たしました。実際に使用してみるとその違いは明らかで、今までの翻訳結果はどうしても不自然な文章とならざるを得なかったのに対し、アップデートされたGoogle翻訳は実に自然な文章の生成を可能にしました。これも元来の辞書ベース、ルールベース、古典的な機械学習手法による翻訳モデルではなく、大量の翻訳文書対を多層再帰型ニューラルネットワークで学習させることにより実現しています。

その実力は音声認識分野でも発揮されています。この分野では2010年代以降、音波の特徴量抽出と文章予測にDeep Learningが使用されるようになり、誤認識率を10%以上下げることになりました。現在では特定の環境下では誤認識率は5%程度まで達成できており、この数字は人間の認識率と同等の実力です。

上述のようにDeep Learningはその深く複雑なネットワークによる学習能力から、ルールが決められたゲームで人間に勝利するにとどまらず、画像や文書、音声の認識などの明確なルール付けが難しく、人間が自身の中にある感覚で認識、判断しているような問題設定に関しても人間と同等の能力を発揮するようになってきており、まさに人工知能としての振る舞いを再現するまでになってきています。

Deep Learningの本質は「データ分析の発展」、さらなる試行錯誤が必要

前節まででDeep Learningの概要と、それがもたらす革新に関して紹介してきました。現在空前のAIブームが訪れており、人間を超えるAIが出てきたり、いずれ多数の職業がAIに取って代わるとの予測がされたりするなど、AIに対する期待が大きく膨らんできている。ビジネスシーンにおいても、「とりあえずAIを使ってみたい」「自社内のデータをDeep Learningさせれば何かすごいことができるか」というように、漠然と期待する方もいるかと思われます。

確かにDeep Learningはデータ活用に革新を起こし、その応用分野も広く、現在も更なる進化を続けています。しかし実際には何でもかんでもDeep Learningをすればいいかというと、そうではない場合が多く、また実際の問題を適切に解くDeep Learningモデルを構築するには高度な数学、統計学、情報工学などの知識が必要となります。一言にDeep Learningとまとめられることが多いですが、問題設定によって行われる処理は全く異なり、前出の囲碁、画像認識(図5)、翻訳(図6)、音声認識でもそれぞれで全く異なった構成のニューラルネットワークが使用されています。例えば画像認識では画像の一部分の特徴量を抽出する畳み込み層と特徴をまとめるプーリング層という役割の層を重ねて用いられます(図4)。翻訳では文書の単語列を時系列データとして処理

するための再帰型ニューラルネットワークが用いられ、膨大な単語空間を符号化、複合化する層と組み合わせられます（図5）。

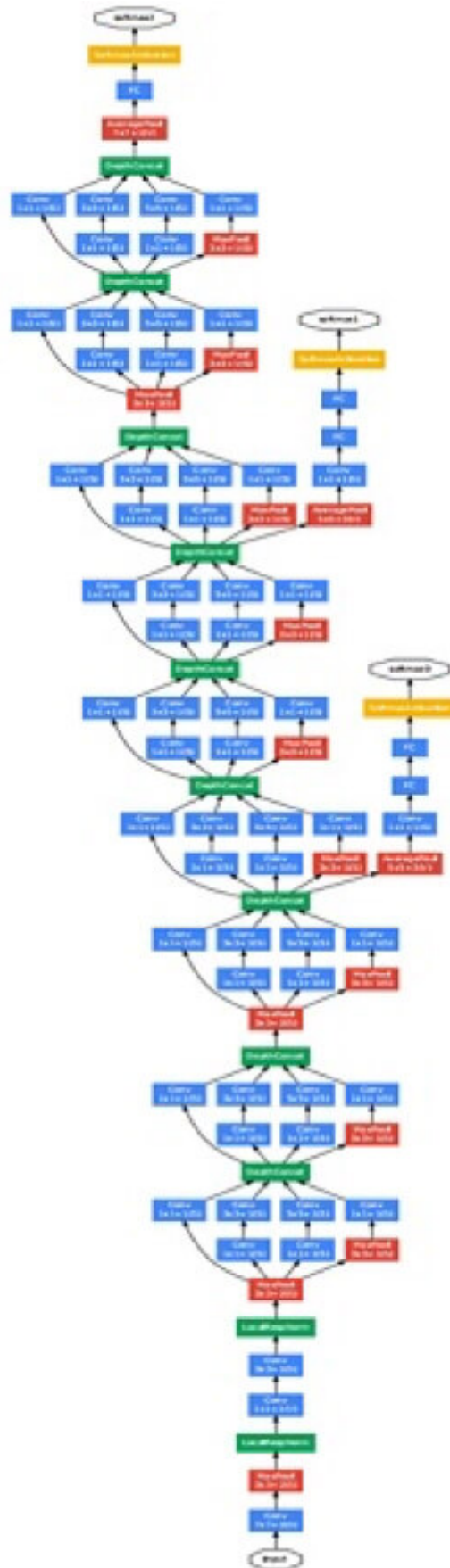


図5 ILSVRC2014の画像分類問題で優勝したGoogLeNetモデル

畳み込み層（青）、プーリング層（赤）、全結合層（黄）と呼ばれる層を複数組み合わせ構成されている。

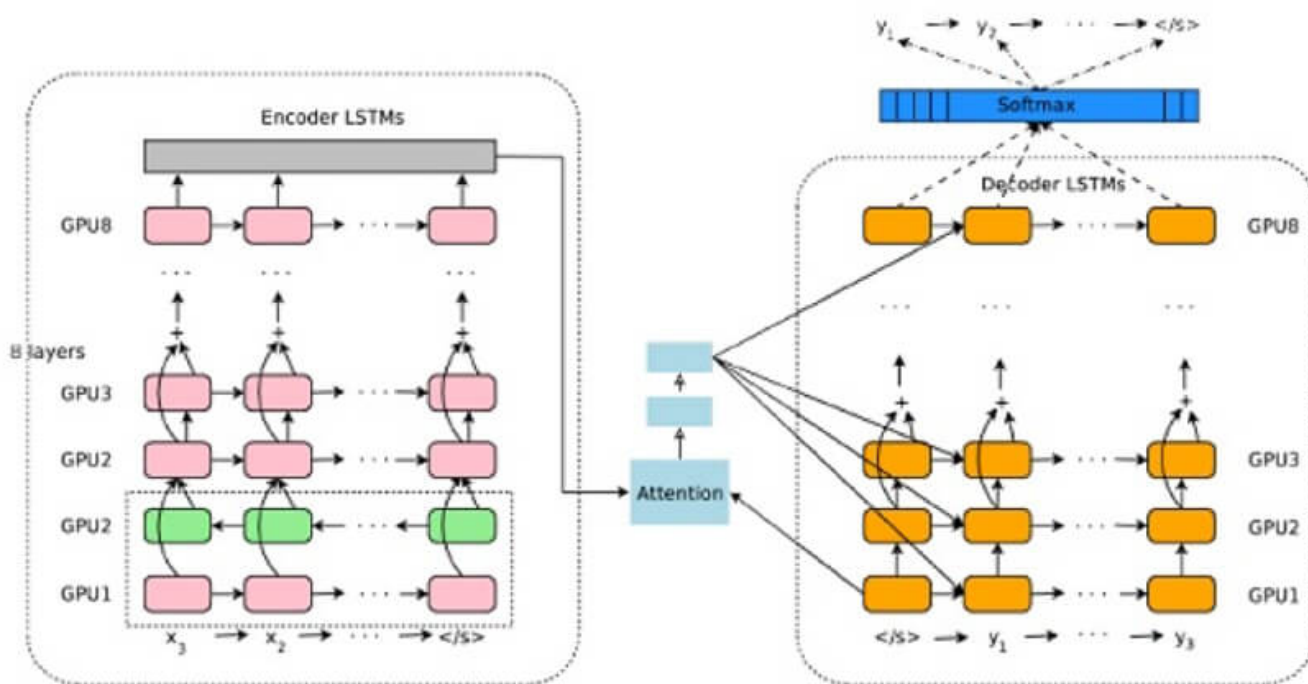


図6 Google翻訳で用いられているネットワークモデル

事前情報を記憶する層、文書の需要箇所を学習する層、次元縮約のための層を組み合わせている。

Deep Learningはただ単に中間層を深くすることで問題を解決可能にしたわけではありません。深く重ねられたニューラルネットの各層には特徴量を抽出する層、次元縮約をする層、事前情報を記憶しておく層、誤差を評価する層など、それぞれ異なった役割があり、その構成やパラメータは入力データと解きたい問題設定により緻密に設計しなくてはならず、そのノウハウは一朝一夕で身につけられるものではありません。とりあえず今あるデータをDeep Learningに通してみようくらいの感覚では良い結果は得られず、適切なネットワークとパラメータを設定できないと、その結果は既存の機械学習手法に劣るものとなります(注)。流行のAIやDeep Learningを活用してみたいという場合、その本質はデータ分析の発展であり、まだまだ我々人間がデータの特性を見ながら試行錯誤しなくてははいけないことを再度認識して取り組んでほしいと思います。

とはいえDeep Learningは今までのコンピュータの可能性を押し広げ、我々に新たな体験を提供してくれており、その可能性は今なお発展の余地を残しています。それは単に機械学習のアルゴリズムの研究分野としての発展だけでなく、それを支えるコンピュータの演算能力の進化とインターネットの発展による学習データの増加にも支えられています。2045年までに技術的特異点(シンギュラリティ)が起こるとの予言もあり、その真偽は別としても、それほどまでにこの分野は今なお目覚ましく発展し続けている。Deep Learningが今後も我々に様々な革新を提供してくれることを期待せずにはられません。

(注) 富士通総研ではDeep Learningを用いた言語処理、画像認識技術の研究開発、またその業務適用事例を多く持ち、経験豊富なデータサイエンティストによるデータ活用視点のコンサルティングサービスを展開している。



知創の杜

あしたを創るキーワード

Deep Learning概説 -AIの核となる機械学習技術の最先端-

当記事の詳細をPDFでご覧いただけます。

[資料ダウンロード](#)

[コンサルタントやエコノミストの知見・ノウハウを紹介する情報誌「知創の杜」はこちら](#)
