

ディープラーニング（米グーグル、米スタンフォード大学、東京大学、産業技術総合研究所）

人工の神経回路、威力増す

「ディープラーニング」と呼ぶ人工知能技術が高い関心を集めている。画像や音声の認識精度が大幅に高まるため、米グーグルなどが研究に参入。経済動向の予測や新薬開発などにも威力を発揮する可能性がある。

ここ1~2年、世界中の人工知能の研究者から大きな注目を浴びている技術がある。コンピューターに人間と同じように経験に基づいた行動をさせる機械学習の一種で、「ディープラーニング」と呼ばれる新手法だ。

インターネット社会を支える画像認

識や音声認識、新薬開発に役立つ化合物の活性予測——。こうした技術の精度を競うコンテストで、ディープラーニングが過去の記録を大幅に塗り替え、次々と優勝を果たしている。

「これほど飛躍的に精度が向上するとは信じられない」「まさに衝撃的な

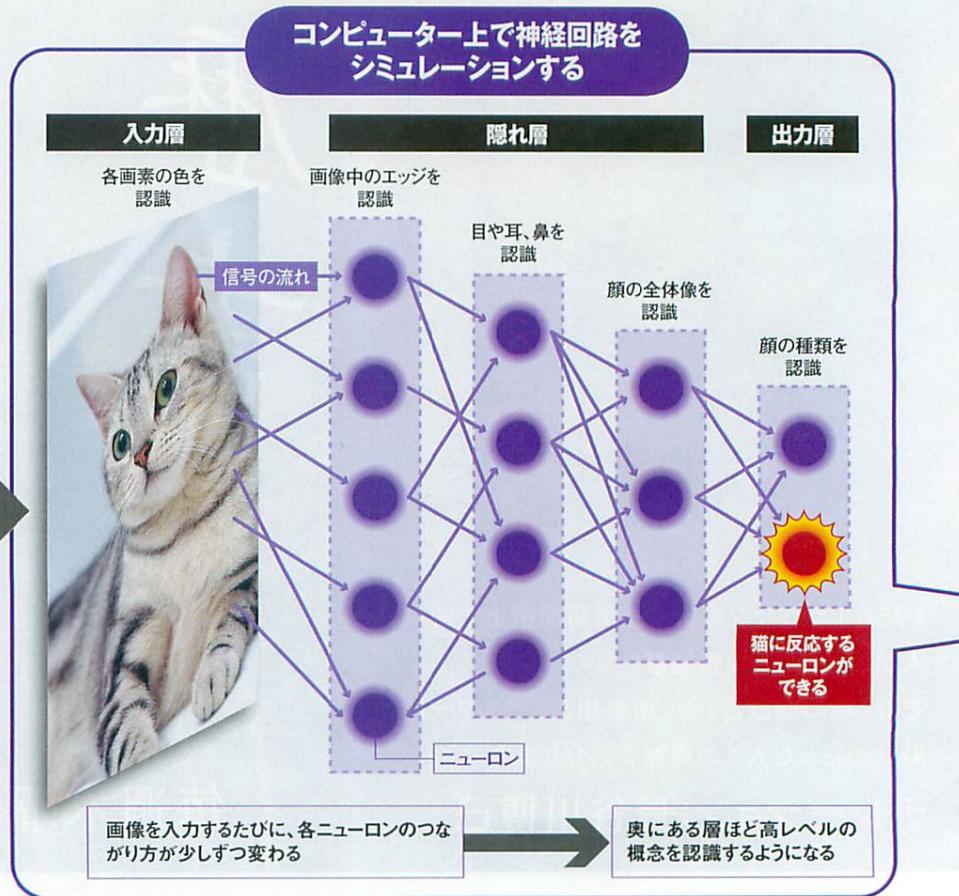
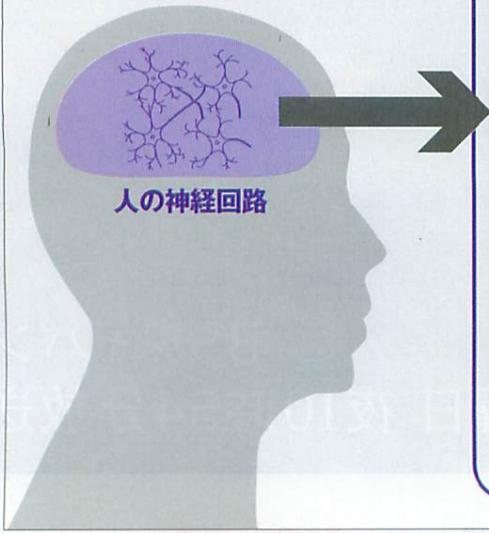
結果だ」。専門家からは、日々驚きの声が上がる。

ディープラーニングは、人の神経回路をコンピューター上で模擬する「ニューラルネットワーク」という技術を発展させたものだ。

人の脳は、画像からそこに映るモノ

ディープラーニングで 画像を認識する流れ

ディープラーニングでは、コンピューター上に人間の脳と同じような多層の神経回路を作製。大量の画像や文字情報を入力してトレーニング(訓練)すると、そこに含まれる高度な概念が自然に引き出される。米グーグルの研究では、出力層のニューロンが、「猫」を認識して強く反応するようになった(右は簡略化した仕組みの図)



の種類を判断したり、手元にある情報から将来のことの予測したりといった能力に優れる。

ニューラルネットは、コンピューター上に仮想のニューロン（神経細胞）を層状に並べ、その間の電気信号の流れを模擬する。これにより、様々なデータから本質的概念を引き出せる、脳の機能を再現することを目指している。

このニューロンの層を4層、5層と“ディープに（深く）”して、脳に迫る性能を引き出そうというのが、ディープラーニングの狙いだ。

ニューラルネット研究の歴史は古く、1950年代にさかのぼる。ただ、ニューロンの層を増やすと性能が落ちるなどの問題を抱え、研究のブームが盛り上がっては下火になることの繰り返しだった。

だが、2000年代の半ばに、米国の研究者がこの課題をクリアする計算技法を開発。大量のデータを処理できるコンピューターの進歩も後押しし、ディープラーニング研究の幕が開けた。

2012年6月、米グーグルが米スタン



米グーグルの研究で、猫を認識するニューロンを最も強く反応させた画像

フォード大学と共同で発表した成果がディープラーニングの威力を世に知らしめるきっかけになった。

グーグルは、動画サイト「YouTube（ユーチューブ）」から取得した大量の画像をニューラルネットに入力。その結果、何も教えていないのに、コンピューターが自然と「猫」という概念を認識した、というのだ。

認識精度が7割向上

グーグルのフェロー、ジェフ・ディーン氏は、「我々の研究は、生まれたばかりの赤ん坊に多数のYouTube画像を見せるのと似ている」と説明する。

同社の研究を例に、ディープラーニングの仕組みを説明しよう。

グーグルは、コンピューター上に合計9層という非常に深いニューロンの層を作った。第1層は「入力層」、第2～8層は「隠れ層」、第9層は「出力層」と呼ばれる。

入力層のニューロンは、入力画像から各画素の値を読み込み、第2層の各ニューロンに信号を伝え。第2層のニューロンは、第1

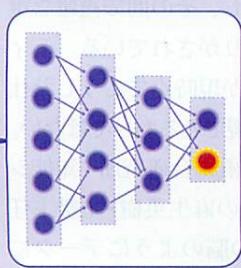
層から渡された信号をある規則に従って変換し、その出力を第3層に受け渡す。第3層以降も、出力層に至るまで同様の手順で出力信号を次の層に渡していく。

ポイントは、1枚1枚の画像を入力するごとに、各ニューロンのつながり方が少しずつ変わるということだ。ニューラルネットに画像を入力すると、各ニューロンが信号を受け取り、変換する際のパラメーター（変数）が調整され、全体的な信号の伝わり方が変わる。いわば、ネットワークが徐々に“トレーニング（訓練）”されていくのだ。

グーグルの研究では、そのパラメー

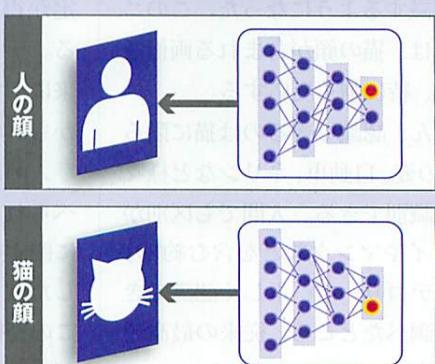
大量の画像で神経回路をトレーニングする

YouTube™



グーグルは1000万本の「YouTube」動画から切り出した画像を使った

様々なモノを自然と認識するようになる



出力層にあるニューロンが、「人の顔」や「猫の顔」など特定の概念に反応するようになる

ネット技術の向上や将来予測に活用 ディープラーニングの応用分野の例

ターが約10億個と、莫大な数に上る。同社は、この巨大なネットワークに1000万本のYouTube動画から切り出した画像を入力。1万6000個のCPU(中央演算処理装置)を結ぶ大規模計算で、1週間がかりでトレーニングした。

トレーニングが終わると、ニューラルネットワークの内部には不思議な構造が出来上がる。各ニューロンが、特定の種類の入力信号に対して特に強く反応(発火)するようになるのだ。

これは、人の脳内で、人の顔や物体など特定のモノや概念に反応してニューロンが発火するのと同じ仕組みだ。つまり、人が学習や経験を通じて作り上げる神経回路の構造が、コンピューター上に生まれたことを意味する。

各ニューロンが何を認識するかは、層の深さで異なる。入力層に近い第2層などのニューロンは、画像中のエッジ(端)やカーブなどの単純な形や概念に反応。一方、第3層、第4層と奥に行くに従い、ニューロンはより高度で抽象的な概念をとらえるようになる。動物の顔の画像を入力した場合、目や耳などの構造や、顔の全体像を認識するのはこうした層だ。

グーグルの研究では、最後の出力層にある1個のニューロンが、猫の顔を自然と認識するようになった。このニューロンは、猫の顔が含まれる画像を見せると、特に強く反応する。

もちろん、認識できるのは猫に限らない。人の姿、自動車、キリンなど様々なモノを識別できる。人間でも区別が難しいエイやマンタなどを含む約2万種類のカテゴリーで、正しく認識できる精度を調べたところ、従来の最高記録を約7割上回る16%に達した。

グーグルは、ディープラーニングの技術を既に同社の音声認識サービスで活用している。今後は、画像認識や動



画像検索



音声翻訳



新薬の効果予測



将来の経済予測

画の検索などの用途でも実用化を検討中だ。グーグルのディーン氏は、「コンピューターと人間との新しいインタラクション(対話操作)が可能になる」と期待を込める。

人間を超える将来予測

人の脳には、千数百億個の神経細胞がある。今後、ディープラーニングの規模や精度が向上すると、人間の脳にどこまで迫れるだろうか。

東京大学大学院工学系研究科の松尾豊准教授は、「人間の脳は、予測や推定が得意なところに最大の強みがある。ディープラーニングを使えば、未来に関する高精度な予測が可能になるかもしれない」と指摘する。

人間は、モノの見た目からそれが食べられるかどうかを判断したり、他人に関する知識からその人の年収を予想したりといった、予測能力に秀でる。この特徴とコンピューターの処理性能が合わされば、人間をしのぐ高精度な将来予測に道が開けることになる。

松尾准教授らは現在、簡易投稿サイト「ツイッター」での発言内容から、投

稿者のプロフィルを性格のレベルまで高精度で推定する研究に取り組んでいる。今後は、スーパーコンピューターなどを使い、ディープラーニングを将来の経済予測に活用する構想も温めている。

ディープラーニングでは、従来技術と異なり、入力するデータをあらかじめ人間が取捨選択する必要がない。このため、新聞に掲載される企業の財務数値や様々な市場のデータ、インターネットの経済情報といった複多な情報を入力するだけで、有用な経済予測が実現するかもしれない。

現実の脳の中には、役割の異なる複数の部位が存在し、その間で複雑な電気信号のやり取りがされている。ディープラーニングが現時点でこうした仕組みをすべて再現できるわけではない。

ただ、産業技術総合研究所・知能システム研究部門の麻生英樹・上級主任研究員は、「人の脳のようにデータに内在する本質的な概念を引き出す技術の開発は非常に重要」と話す。応用対象となる分野も広く、研究の盛り上がりは当分続きそうだ。(田中 深一郎)■