

元データが少ないと AI は学習できないのか？

AI

多様化する学習方法

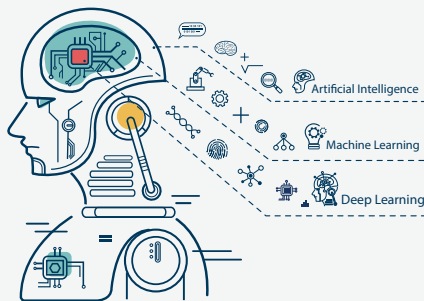
教師あり学習

教師なし学習

… GAN（敵対的生成ネットワーク）が流行

強化学習

… マシン同士で強化しあうことが可能へ



**AI が自ら学習データを生成
することが可能となってきた**

ディープラーニングを行なうにはビッグデータが必要ですが、元にする学習データが大量にないと AI を利用することができないのでしょうか？

もちろん機械学習ができることが AI の前提条件ではないため、大量の学習データが必要かどうかは対象とする業務内容によって状況が異なることは確かです。しかしながら、企業の多くは、人に頼っていた経験による判断や、今後の需要予測のために AI を業務で使用する機会が多いので、そのほとんどは機械学習を使った AI を期待しているでしょう。

その機械学習にも様々な手法があり、一般的に想像し易い学習データに正解ラベルを付ける「教師あり学習」に対し、クラスタリングなどの手法利用した「教師なし学習」もあります。但し、これは、あらかじめ出力すべきものを決めておかないというだけで、学習するデータが少なくてもよいということには関係ありません。

その「教師無し学習」の中で、GAN（Generative Adversarial Networks：敵対的生成ネットワーク）が、実在しないアイドルの画像を生成したり、フェイク動画を作ったりしたことで、近年非常に注目されてきています。

GAN は、競合する 2 つのニューラルネットワークのシステムによって実装され、相手を騙すためのデータを作るためのモデルと、それが正しいかどうかを判断するためのモデルを使ってトレーニングすることで元のサンプルデータが少なくても、マシン上で似たデータを大量に生成することができます。

また、囲碁の世界で、人間を負かして有名になった Google DeepMind のコンピュータ囲碁プログラム「AlphaGo」を破ったのは、ビッグデータ不要で自己対局のみで人工知能同士の対戦という「強化学習」をした「AlphaGo Zero」だったことは、ヒトが学習データをすべて与えなくても AI は学習できることを象徴しています。

また AI は、クラウドのサービスを中心にすでに学習したことを元に、すべてゼロから学習させなくてもよい環境が、整ってきました。これは、ある程度業種や業務形態を特定して、すでに標準的なデータセットの学習を終えた AI をサービスとして提供されています。企業にとって、満足のいく学習ができているのかどうかは確認が必要ですが、業種や業務内容が同じであればすでに学習済みの AI を利用することで、学習させる時間とコストを節約できるようになってきました。これ以外にも、いかに少ない学習データで AI を活用できるようにするかについては、各メーカーがいろいろなソリューションを展開し始めているので、データが少ないことで AI の取り組みを諦めることは必要ない時代になってきています。