

# 特に注目的人工知能 アルゴリズム百科

佐藤 聖

## ● アルゴリズムを組み合わせる進化中

第2章の冒頭で述べたように、個人でも機械学習を試せるハードウェア環境が整いつつあります。これまで単純な機械学習アルゴリズムでは達成できなかった目標も、より高度なアルゴリズムを使えば達成できるかもしれません。

複数のアルゴリズムを組み合わせるアンサンブル学習には、

- 半教師あり学習
- トランスダクション学習
- マルチタスク学習

があります(後述)。それぞれ、さまざまな工夫によって派生的な学習方法が開発されており、いろいろな名前前で呼ばれています。これらのより進んだアルゴリズムによって万能AIを目指す研究も進んでいます。

例えば自動車の自動運転やドローンの自動航行など、達成目標に応じて複数の機械学習アルゴリズム(教師あり学習、教師なし学習、強化学習など)を組み合わせるので、複雑な構造になります。これらの学習方法は、まだまだ研究開発の域を出ておらず、企業や組織での活用例がほとんどありません。ですが今後はコンピュータ性能がさらに向上し、演算コストが下がるので、一般企業でも活用が広がると思います。

機械学習はコンピュータの中だけでなく、IoTやロボットという形で、人間と同じ環境で動き、働き始めています。今後は機械学習のアルゴリズムだけでなく、モータや電子部品も進化するため、より複雑なタスクを簡単にこなせるようになるでしょう。機械学習からますます目が離せませんね。

## 1 半教師あり学習

### ● ビッグ・データのラベル付けが可能に

パブリック・クラウドを利用することで簡単に大規模なサービスを提供でき、そこから収集される膨大なデータが手に入るようになりました。こうしたビッグ・データを持っていたとしても、データベースで処理するには膨大なデータを処理できるコンピュータ性能が必要になります。データのラベル付けを手作業で行ったのでは高コストになるという課題があります。コスト効果を考慮するとビッグ・データの全てにアノテーション作業することは現実的ではありません。

半教師あり学習では、教師なし学習とディープ・ラーニングの教師あり学習と組み合わせるとビッグ・データの処理に効果を発揮します。膨大なデータには必ずしもラベルが付与されていないかもしれませんが、一部のデータにアノテーションできれば作業コストを低減できます。

ビッグ・データの数%にアノテーションして学習し、残りのデータを学習モデルを使ってアノテーション作業しながら教師あり学習をします。膨大なラベル

なしのデータがあったとしても複数のアルゴリズムを組み合わせることでアノテーション作業のコスト削減ができ、ビッグ・データを有効に活用できます。

### ● 教師あり学習の課題を解決する注目株

近年、急速に進化しているアルゴリズムの1つに半教師あり学習があります。

#### ▶ 従来の教師あり学習

CNNやRNNなどの機械学習アルゴリズムで教師あり学習を行うためには、分類モデル構築のためにラベル付けされたトレーニング・データやテスト・データが必要になります[図1(a)]。

ラベルなしデータにラベルを付けるアノテーション作業は人的作業になり、人工知能システム開発で最も大変な作業です。作業を簡単にするアノテーション・ツール(LabelMe, LabelImg, VoTTなど)もたくさんありますが、根本的には手作業ですので、それなりに時間がかかります。自動化して人手を掛けないようにすることもある程度可能です。図1(a)の教師あり