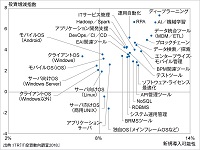
**AIブームを支える「機械学習」～AIの現実的な始め方とは？**

**[佐伯 竜輔](https://enterprisezine.jp/author/499)**[著]

2018/07/18 06:00

　AIブームは日本国内でも着実に広まってきています。あらゆる企業で人手不足が叫ばれている現状を考えると、AIに頼っていくという流れは加速していくでしょう。しかし現状はAIというキーワードと先進的な事例が先行し、その裏側の仕組みや、現実的な取り組み方についての理解が追い付いていないように見受けられます。今回はAIを支えるディープラーニングと機械学習技術の紹介を中心に、導入までに必要な具体的なステップを見ていきたいと思います。また、後半では機械学習など様々な分析を行えるデータベース「Vertica」を用 いて、センサーデータから機械の 異常予兆を検知する方法もご紹介します。



　Verticaの基礎を知りたい方は、アシストが開設している「Vertica技術情報サイト」の記事をご参照ください。**（**[**Verticaとは**](http://vertica-tech.ashisuto.co.jp/about-vertica)**）**

**AIブームの背景**

　最近、「AI」という単語を見聞きする機会が多くなりました。その波はIT業界に留まらず、テレビでも特集が組まれたり、CMで商品のキーワードとして用いられるなど、世の中全体のブームと感じている方も多いのではないでしょうか。ITRの「IT投資動向調査2018」を見ても、AI／機械学習、そしてディープラーニングが新規投資対象分野として大きな注目を浴びていることがうかがえます。

　これまでにも推論・探索やIF～THENベースのエキスパートシステムを利用した小さなAIブームがありましたが、実用的でなかったり、実際に運用するには多くの人手が必要だったりと、AIという単語から連想される、まるでコンピュータ自身が考えているような振る舞いには到底及びませんでした。

　しかし、2012年にGoogleからコンピュータに大量の猫の画像を与えることで、コンピュータ自身が猫の特徴を理解し、猫の画像を認識できるようになるという研究結果が発表されました。この裏側で利用されている技術がディープラーニングです。また、2015年にはGoogle傘下のDeepMind社が開発したAlphaGoという囲碁AIが人間のプロ囲碁棋士に勝利しました。このAlphaGoにもディープラーニングが使われていました。AIが人間のプロ棋士に勝利するにはあと10年はかかるだろうと言われている中での勝利はとても大きな話題になりました。

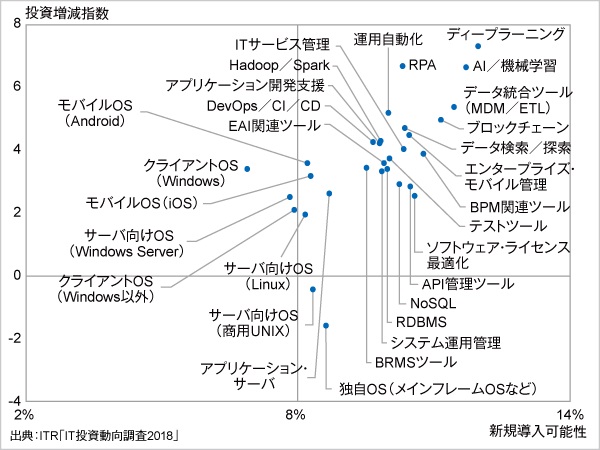


図1　OS/ミドルウェア分野への投資意欲（プロット図）  
（出典：ITR「IT投資動向調査2018」）

　このようなAIにとっての明るい話題が重なったことでディープラーニング技術に注目が集まり、昨今のAIブームに繋がっていきました。

**ディープラーニングのハードル**

　前述のとおり、現在のAIブームのきっかけはディープラーニングの登場によるものといえます。そのため、企業がAIを導入しようとした場合「ディープラーニングを使って課題を解決したい」という考えから始めるケースが多いようです。しかし、ディープラーニングの導入は単純ではなく、複雑なパラメータチューニングが必要であったり、学習のために大量のデータが必要だったりと、導入へのハードルは高いと言えます。ではこれらをクリアできない企業はAIの導入を諦めなければいけないのでしょうか。実は解決したい課題によっては、ディープラーニングが登場する前から利用されているクラシックな機械学習で十分に対応できるケースが意外と多くあります。

**ディープラーニングと機械学習の違い**

　ここでディープラーニングと機械学習について整理してみましょう。

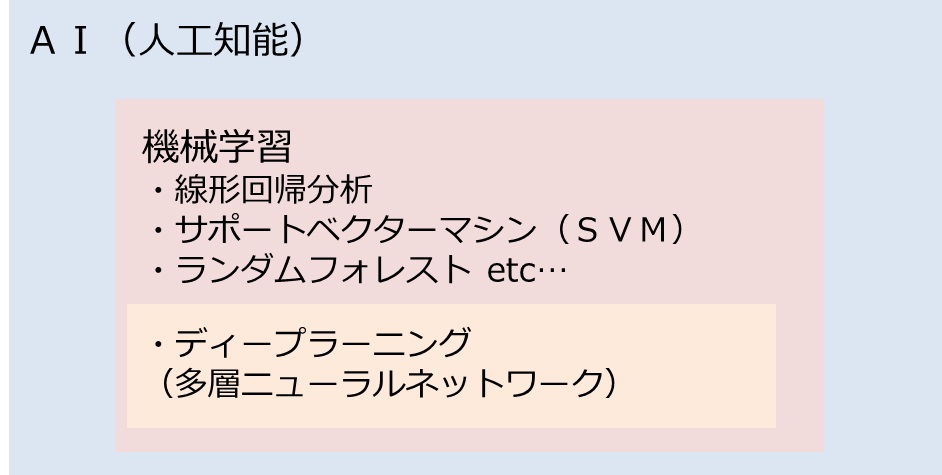


　図2にあるようにディープラーニングは機械学習の中の一つの手法に過ぎません。機械学習にはディープラーニング以外にも統計解析でも利用される線形回帰分析や、今でも根強い人気を持つサポートベクターマシン（SVM）といった様々な手法があります。これらの機械学習をディープラーニングと区別するために、「クラシカルな機械学習」や「旧世代の機械学習」「統計的機械学習」といった呼び方がされることがあります。

　ディープラーニングとクラシカルな機械学習はどちらも「何かを予測する」ために利用しますが、最大の違いは予測に関係がある情報（特徴量）を人がコンピュータに教える必要があるか、コンピュータ自身が見つけることができるかです。例えば赤りんごと青りんごを見分けるAIを作りたい場合、クラシカルな機械学習では特徴量として、りんごの形、糖度、色といったりんごに関する様々な情報の中から関連がありそうな特徴量を人が選んでどこに着目すべきかコンピュータに教える必要があります。

　これに対してディープラーニングは特徴量を自分で見つけることができます。つまり赤りんごと青りんごを見分けるために「色」に着目すべきということを自身で見つけることができます。しかし、ディープラーニングを用いた予測結果は「なぜそういう予測に至ったのか」という理由を知ることが難しいというデメリットもあります。一方クラシカルな機械学習は確立された統計的手法がベースになっており、ディープラーニングと比較すると予測結果に対して理論的な根拠を得られやすい傾向があります。

## 機械学習の基本４ステップ

　機械学習を利用する場合、一般的には次の四つのステップを通る必要があります。

##### ①データ準備

##### ②予測モデル作成

##### ③評価

##### ④実装

　ここからは機械学習で必要になる四つのステップの詳しい内容を見ていきましょう。

##### ①データ準備

　まずはコンピュータに学習をさせるためのデータを準備する必要があります。ここでは予測したい事象に関連しそうなデータを社内のデータベースなどから集めてきます。また社内だけでなく社外のSNSデータやオープンデータと組み合わせて利用することもあります。データを集めたらすぐに機械学習ができるかというとそうではなく、機械学習のために文字データを数値データに変換したり、欠損値の補完や異常値の排除といった前処理が必要なケースがほとんどです。このデータ準備フェーズは機械学習全体の8割を占めると言われることもあります。この分野は「データプレパレーション」と呼ばれ注目を集めており、アシストでもデータプレパレーションツール「Paxata」を投入することで、市場のニーズに対応しています。

##### ②予測モデル作成

　データ準備が完了したら、そのデータを使用して予測モデルを作成します。このステップではサポートベクターマシン等のアルゴリズムの利用を検討したり、利用する場合はアルゴリズムのパラメータチューニングを行います。

##### ③評価

　予測モデルを作成したら、そのモデルで本当に正しい予測ができるのか評価（テスト）します。多くの場合、1回目の評価では十分な予測精度を出すことはできず、再びデータ準備フェーズに戻って、使用するデータを変えてみたり、アルゴリズムを変更するなど、試行錯誤を繰り返すことになります。

##### ④実装

　十分な予測精度を持つモデルができあがったらいよいよ実装です。実装は要件に合わせて様々な方法が考えられます。例えば、予測結果をBIツールを使ってリアルタイムに確認したい場合は、BIツールと予測を行うアプリケーション（作成した予測モデルが存在するサーバ）を連携させ、バッチ処理で構わない場合は、夜間にバッチ処理で予測を行い、その結果をBIツールがアクセスするデータベースにロードしておくといった方法等が考えられます。

## どこで機械学習を実施するか

　機械学習をするための代表的なプラットフォームは二つあります。一つはクライアントPCです。機械学習を行うための言語はRまたはPythonが使用されることが多く、これらの言語はクライアントPCでも簡単に動作させることができます。気軽に利用できる反面、クライアントPCで機械学習を行うにはマシンスペックの限界があるため、大量データの場合はある程度データのサンプリングを行う必要があります。また、機械学習の予測結果を社内のアプリケーションと連携させたい場合もクライアントPCで作成したモデルを予測を行うサーバ上に移植する必要が出てきます。

　二つ目として、もう少し大規模に機械学習を行いたい場合は、Hadoop ／Sparkを利用する方法があります。特にSparkはインメモリで高速な分散処理が行えることに加え、Spark MLという機械学習のためのライブラリが用意されているため、Spark上での機械学習が実施しやすくなっています。

## データベースで機械学習をするという新たな選択

　アシストでは高速検索・分析を得意とするデータベース「Vertica」を取り扱っており、高速な検索・分析が求められるデータウェアハウス（以下、DWH）として採用いただくケースが増えています。VerticaはDWH用途以外にも、前述した機械学習に必要な1～4のステップ全てが行える機能を備えていて、ビッグデータ分析基盤としても利用可能です。そのため、Verticaを導入すれば、追加投資なしで機械学習にチャレンジすることができます。例えば、センサーデータを学習する場合、①データ準備では「欠損値補完」、「一定時間単位の集計」、「ギャップ補完」といった時系列データ処理が必要ですが、Verticaにはそれらを処理するための関数が用意されています。

　また、④実装ではVerticaで作成した予測モデルはSQLで呼び出すことができるので、アプリケーションとの連携が容易といった利便性もあります。

## IoT×機械学習でリアルタイム予測に挑戦

　AIを利用した「予測」に対するニーズはますます高まっています。最近ではIoTデータと組み合わせて、異常検知などを行い、それらを素早く意思決定に活用したいというリアルタイム性も強く求められています。例えば、製造業界では機械の異常を機械学習で予知し、壊れる前にパーツ交換等の対処を行う「予知保全」という考え方が注目を集めています。

　アシストでは昨年、VerticaとBI製品であるWebFOCUSを組み合わせたIoT×機械学習のリアルタイム故障予測デモンストレーションを作成しました。このデモンストレーションではモーターで駆動する装置に振動を検知するセンサーを取り付け、それを工場の機械に見立てています。発生したセンサーデータをリアルタイムにVerticaにロードして、そのデータをVertica内の機械学習機能で分析します。意図的にモーターの電圧異常を引き起こすと、通常とは異なる振動が発生し、Verticaが異常が発生したとの予測結果を算出します。この予測結果をWebFOCUSで呼び出すことで機械の状態をリアルタイムに知ることができます。

　もし異常と判断された場合は、WebFOCUS上で詳細分析して、その機械が故障することでどれくらい製造ラインの稼働に影響が出るのかといったことが確認できます。そして、もし影響が大きければ、スケジュールを調整をして緊急メンテナンスを行うといった具体的なアクションにつなげることができます。これにより、発生していたかもしれない損害を未然に防ぐ予知保全を実現できます。このデモンストレーション動画を[**YouTubeで公開**](https://youtu.be/ZyPZgCJngWI)していますので、ぜひご覧ください。

　また、Verticaで機械学習を行うメリット等をまとめた「Vertica紹介資料セット」も[**こちらのサイト**](http://www.ashisuto.co.jp/product/category/database/vertica/download/#tab)からダウンロードいただけます。

　「AI」という言葉だけを見ると、一部の先進的なユーザーだけのものと感じるかもしれませんが、小さく始めようと思えばご紹介したようにBIとデータベースの組み合わせだけでも実現することは可能です。アシストは今後もAIやIoTなどの最新技術を実際にお客様に活用いただくための方法を提案・模索して参ります。

##### 参考文献：『人工知能は人間を超えるか ディープラーニングの先にあるもの』（角川EPUB選書、松尾 豊 著）

＊＊＊  
\*本記事はアシスト広報誌『アシスト』No.31　2018年春号に掲載されたものを一部加筆・修正して転載しています。