

# 脱！ブラックボックス 説明可能なAI

新連載

AIモデルが重視している特徴を抽出…  
第1回 線形回帰モデルの場合

切通 恵介

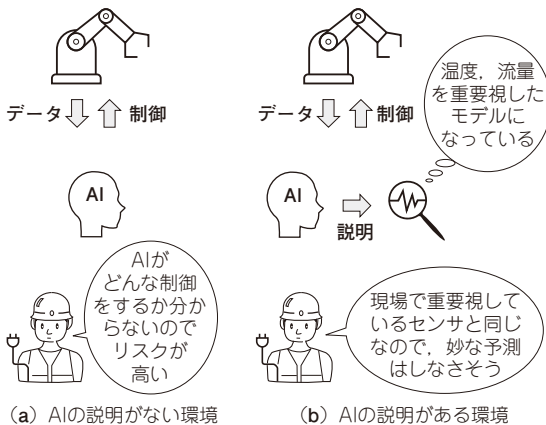
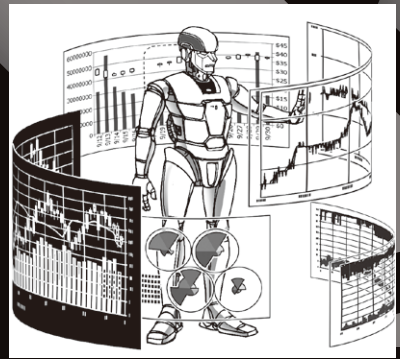


図1 AIモデルが重要視しているパラメータが妥当なものなのか分かって安心

## ● 連載概要

ディープ・ラーニングや機械学習などのAIモデルは、複雑性が高くAIモデルの中で何が起きているのかは、多くの場合、ブラックボックスとして扱われます。このため、AIが出した結果の根拠を出すのが難しく、産業への応用には課題があります。ここでも

し、AIが重要視した特徴が、人間の専門知識に基づく判断と一致していれば、その結果も妥当であると判断できる可能性がでてくるのではないのでしょうか(図1)。

そんな中で説明可能なAI(eXplainableAI: XAI)という概念が生まれました。経済産業省のガイドラインの中でも言及されていて、ビジネスにおいて重要な概念となっています。

説明可能とは、AIモデルの説明が得られる、すなわち、AIモデルが出した結果の根拠を示せるということです。結果の根拠とは、モデルへの入力特徴の重要度を数値化したものや、計算過程を可視化したものです。現在、さまざまな手法が提案されています。本連載では、AIモデルが出した結果の根拠を計算するアルゴリズムとその実装までを解説します。

## ● 今回のテーマ

今回は説明可能なAIが求められる背景と、最も単純なAIモデルである線形回帰モデルの場合について、気温などの簡単な数値を使った実験と、ワインの品質データを使った実験を例に、AIモデルの「説明」とは具体的にどんなものなのか解説します。

## 産業界で説明可能なAIが求められる背景

### ● AIモデルの高機能化、複雑化でますますブラックボックス!

AIモデルは日々進歩しつづけています。画像認識や文章生成などの分野で用いられるディープ・ラーニング・モデルや、Kaggleなどのコンペでベースラインとしても用いられるLightGBMなど、複雑なモデルが高い精度を達成し、ビジネス課題を解決できるようになりました。一方で、モデルが複雑になればなるほど、モデルがブラックボックス化されてしまい、その結果の解釈が難しくなるという課題もあります。これは特にAIモデルによる予測や判断が重要になる分野では致命的とも言えます。

### ● AIモデルの判断が妥当か知りたい

例えば、製造業においてAIモデルの予測値に基づいて自動でプラントを制御するような場合を考えます(図1)。

このAIモデルは複数のセンサを入力として、制御値を予測値として出力するとします。もともとは人がセンサを見ながら制御していたもので、2年分のデータで学習、検証を実施したところ、予測値と理想的な制御値との平均誤差が人間が制御した場合を上回り、机上のデータでは歩留まりが改善されるという指標が得られました。さて、このモデルを現場導入する決断をできるでしょうか。