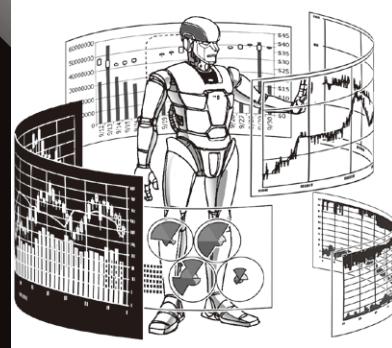


脱! ブラックボックス 説明可能なAI

第5回 モデルの学習が適切か判断できる!
モデルに依存しない説明手法…PFI



切通 恵介

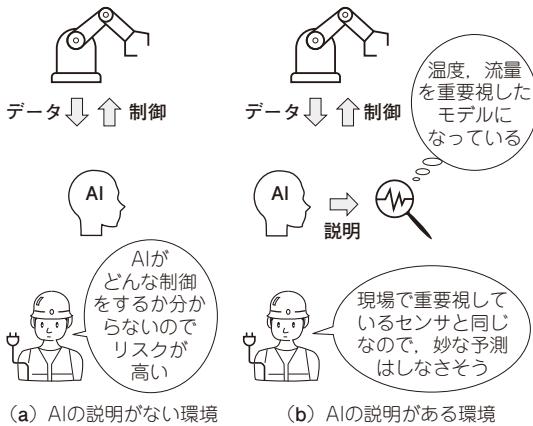


図1 AIの説明性有無による違い

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。そんな中で説明可能なAI(eXplainable AI:XAI)という概念が生まれました(図1)。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。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

今回はモデルに依存しない説明手法であるPermutation Feature Importance(以下、PFI)を解説

します。この手法はどのような機械学習モデルに対しても説明できる点は前回紹介したPartial Dependence Plot(以下、PDPlot)と同じです。PDPlotは、注目している特徴量を変化させたときのモデルの出力に注目した手法であるのに対し、今回説明するPFIは、モデルの予測誤差からモデルの振る舞いを計算する手法だということです。

AIを説明する手法はいろいろある

● 説明手法は4つに分類できる

説明可能なAIの要素技術としてAIの振る舞い、すなわちAIの説明を抽出する技術が開発されています。これらはGlobal(大局的)かLocal(局所的)かの軸、Model specific(モデル依存)かModel agnostic(モデル非依存)かの軸で大きく4種類に分けられます(表1)。本連載でもこれらの軸で大きく4種類に分かれながら紹介します。

● Model specific(モデル依存)な説明の限界

実際のビジネスにおいて機械学習を用いた課題解決を実施する際は、複数種類のモデルを利用し、その精度や説明を比較することができます。

例えば、天気や降水量などの天気予報データから、ある商品が1週間後にどれぐらい売れるのかの需要予測を線形回帰モデルと勾配ブースティング木で行うとします。そのAIの説明においてModel specificな手法、すなわち線形回帰の重みの可視化(第1回、2025年3月号)と、どの特徴量が結果に貢献しているか示すFeature Importance(第3回、2025年7月号)を利

表1 説明手法の分類

説明のタイプ	Model specific(モデル依存)	Model agnostic(モデル非依存)
大局的	<ul style="list-style-type: none"> 線形回帰の重み(連載第1回) 決定木の構造の可視化(連載第2回) Feature Importance(連載第3回) 	<ul style="list-style-type: none"> Partial Dependence Plot(連載第4回) Permutation Feature Importance(今回のテーマ)
局所的	<ul style="list-style-type: none"> ニューラル・ネットワークの勾配(Saliency Map, Integrated Gradients) 	<ul style="list-style-type: none"> LIME(Local Interpretable Model-Agnostic Explanations) SHAP(Shapley Additive Explanations), Kernel SHAP Anchors

第2回 決定木回帰モデルの判断が妥当か調べる(2025年5月号)

第3回 どの特徴量が貢献しているかを表すFeature Importance(2025年7月号)

第4回 入力特徴量が多くてもOK! モデルに依存しない説明手法…PDPlot(2025年10月号)