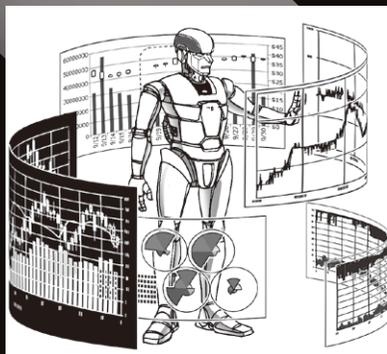


脱！ブラックボックス 説明可能なAI



第2回 決定木回帰モデルの判断が妥当か調べる

切通 恵介

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示すのが難しく、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。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

第2回目は決定木モデルに着目し、その分割方法のアルゴリズムの解説と、モデルの説明としての分割構造の可視化をPythonコードを用いて実装します。

● AIを説明する手法にはいろいろある

AIの振る舞い、すなわちAIの説明を抽出する技術は、表1のように大きく4種類に分けられます。本連載ではAIの説明の手法について4種類に分けて紹介します。

表1 説明手法の分類

説明のタイプ	モデル依存	モデル非依存
大局的	<ul style="list-style-type: none"> 線形回帰の重み 決定木の構造の可視化 Feature Importance 	<ul style="list-style-type: none"> Partial Dependence Plot Permutation Feature Impotence
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

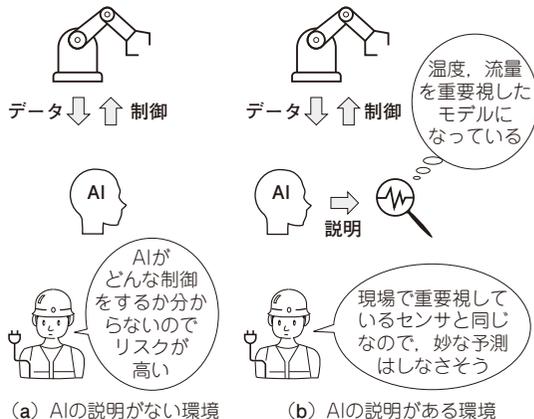


図1 AIの説明性

▶ Global (大局的) な説明か Local (局所的) な説明か

大局的な説明手法としては、線形回帰モデルなど機械学習モデルの重みを可視化するものがあります。モデルの重みはデータ・サンプルによって変化しないため、全ての入力サンプルに対して同じ説明が得られます。一方、局所的な説明とは、注目しているサンプルに対して各特徴の重要度を計算するものです。例えば、Saliencymapという方法では、画像を入力したときに、その画像のどのピクセルをどれくらい重要視したのかの値を抽出できます。こういった手法は画像が変わるごとにその値が変化します。