

AI判断の根拠を説明する XAI を使いこなす

編集にあたって

袖美樹子 | 国際高等専門学校

佐々木貴之 | 横浜国立大学

AI と言えば深層学習と言ってよいほど深層学習が現場で導入されるようになってきている。深層学習の問題はなぜその結果が導かれたのか説明性に欠ける点にある。結果は正しそうだが本当に正しいのか？なぜその結果が導かれたのか？ほかの解はないのか？人は理解し納得して使いたい。その要求に答える技術が XAI (eXplainable AI) である。

2016 年に予測の根拠となった特徴量を提示する LIME が発表されると XAI の実用化研究が盛んに行われるようになった。現在では各種シーンで導入が進められている。XAI の導入にあたってはいろいろな障壁がある。最も大きな障壁は、誰に説明をするかにより説明方法が異なる点である。技術者と一般利用者では要求される説明内容に差がある。しかし XAI は人に合わせた説明を行えるわけではなく、システムを提供する人が利用者のニーズを掴み分かるかたちに説明を変更し提示を行う必要がある。すなわち現在 XAI を使いこなすには利用形態を理解しカスタマイズをする必要があり、この部分が技術者の方の見せ所である。

本特集ではまず XAI がどのような技術なのかを解説いただく。次に利用に際し必要となる勘所が分かるよう IT を得意とする企業の第一線で活躍する技術者の方々に導入事例を紹介いただくこととした。

XAI を使いこなす技を読み取っていただければと思う。

第 1 の記事は、「説明可能 AI (XAI) とは？～深層学習の説明性向上と XAI の今後の展望～」と題して横浜国立大学長尾智晴先生に XAI の技術概要を解説いただいた。何をもって説明 (納得) の拠り所にするのか？ XAI 実用化の鍵であり、永遠の課題でもある。本記事では実際の例を用いながら AI の判断根拠の考え方、技術の根幹が理解できるよう解説いただいた。また、今後人と機械が互いに相補的に知能を高め合い共に進化する社会実現の可能性を解説いただいた。

第 2 の記事は「産業利用における説明可能 AI の使いどころ」と題して (株) NTT データ 坂元哲平氏、安部裕之氏に AI の産業利用における課題と解決策を解説いただいた。AI の産業利用を促進する技術として XAI が提案されているが、技術だけでは課題解決に至らない場合もあり、XAI の適切な使い方の整理が重要になっている。本記事では、XAI 技術の使い方の側面から解説いただいた。

第 3 の記事は「制御の根拠を明示できる XAI の取り組み—DX × UI による AI 説明性向上技術—」と題して三菱電機 (株) 毬山利貞氏、横須賀佑介氏、穂苅寛光氏にオフィス内空調制御への活用方法につ

いてエンドユーザが安心して AI を利用いただくための取り組みを解説いただいた。エンドユーザにとっては、AI の今後の振る舞いが予測可能であることやなぜそのような振る舞いをしたのか理解できない場合その原因を調べられ、対処できることが安心につながる。エンドユーザに安心を与える取り組みを解説いただいた。

第4の記事は「Shapelets 学習によるインフラ・製造分野向け時系列波形の異常診断技術—異常の検知や診断に有効な波形パターンを発見する AI—」と題して(株)東芝研究開発センター 山口晃広氏にインフラ・製造分野における異常診断向け XAI 技術を解説いただいた。インフラ・製造分野では異常の原因究明には AI の判断根拠を分かりやすく専門家に提示する必要がある。また時系列データの事前学習なしで異常検知する技術も重要である。本記事ではこれらの技術への取り組みを解説いただいた。

第5の記事は「説明可能な AI を身近にするためのディープラーニングツール」と題してソニーグループ(株) R&D センター 鈴木健二氏に AI の判断根拠を可視化させ、AI の認識精度を向上させるディープラーニングツールについて解説いただいた。解説いただくツールは専門的なスキルやプログラミング技術を必要とせず直観的に利用できる。XAI を使いこなすのは敷居が高いが、ツールを活用することにより敷居を下げるのが可能となる。

第6の記事は「信頼できる AI の実現に向けて

— XAI による根拠の納得感向上のアプローチ—」と題して(株)日立製作所研究開発グループ 恵木正史氏、間瀬正啓氏、濱本真生氏に Shapley 値を用いた根拠説明手法の考え方を解説いただいた。現場から信頼される AI を実現する上で重要な論点の解説である。

SF アニメ作品『新世紀エヴァンゲリオン』では「マギ (MAGI)」という架空のスーパーコンピューター・システムが登場する。カスパー (Casper)・メルキオール (Melchior)・バルタザール (Balthasar) の3台からなり、それぞれに赤木ナオコ博士の「女性の部分」「科学者の部分」「母の部分」の思考パターンを持った巨大コンピューターシステムである。3台の思考パターン (プログラム) をそれぞれ微妙に変えることで、人の持つジレンマが再現されていると言われている。

AI が広く実際の現場で使われるようになり XAI の実用化が進んでいることは画期的だ。XAI の挙動を観察すると同じ問題を与えても用いる AI 技術によって結果が異なる。またその説明根拠も異なる。これらを相互に学習し技術を磨けば精度は上がっていくに違いないが、究極の進化を遂げた後でも「マギ」のように人格として根差す考え方により答えやそれを導いた根拠は異なりそれは解消できないのだと思う。より人間に近づいたということではあるが、今後も AI 技術の進化から目が離せそうもない。

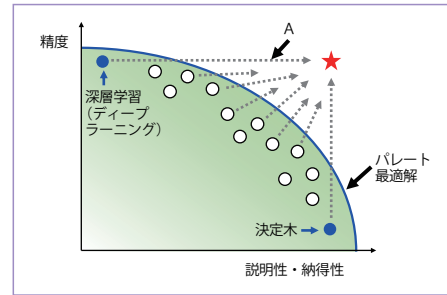
(2022年6月10日)

概要

1 説明可能 AI (XAI) とは? ～深層学習の説明性向上と XAI の今後の展望～

長尾智晴 | 横浜国立大学

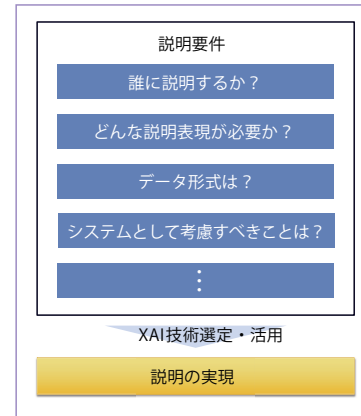
本稿では“説明可能 AI (XAI: eXplainable AI)”の概要について解説している。機械学習の精度と説明性の関係に触れた後、説明性の定義と、XAI システム作成上の注意、企業での AI 利用、深層学習の説明性を高めるためのさまざまな手法の概要を紹介している。最後に、XAI から共進化型 AI (CAI: Co-evolutional AI) への移行と XAI の将来展望についても述べている。



2 産業利用における 説明可能 AI の使いどころ

坂元哲平 安部裕之 | (株) NTT データ

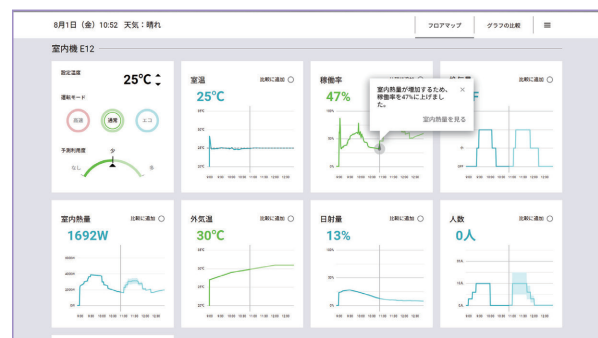
近年、AI の産業利用にあたって、AI の説明性が課題となっている。この課題を解決する技術として、説明可能な AI (eXplainable AI, XAI) が提案され、多くの技術が報告されている。一方、技術だけでは課題の解決に至らない場合もあり、XAI の適切な使い方の整理も重要になっている。本稿では、XAI を利用するデータ形式・システム・人の面から、技術の使い方や留意点、課題について解説する。



3 制御の根拠を明示できる XAI の取り組み — DX × UI による AI 説明性向上技術 —

毬山利貞 | 三菱電機 (株) 情報技術総合研究所 横須賀佑介 | 三菱電機 (株) 統合デザイン研究所
穂苅寛光 | 三菱電機 (株) 情報技術総合研究所

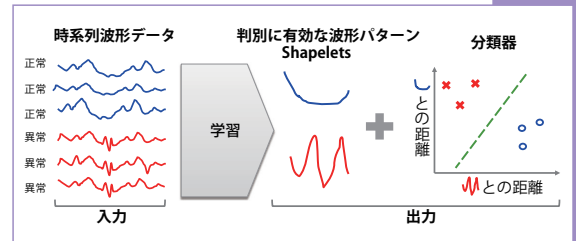
AI の説明能力の向上を求める動きが世界的に加速している。本稿では空調制御を例とし、DX と UI を組み合わせることで、機器制御に対する AI の説明性を向上させる取り組みを紹介する。紹介技術では、室内環境のグレイボックスモデルを用いて環境変化を予測し、ユーザに AI の今後の振る舞いを可視化する。振る舞いに変化が発生した場合にはその変化の原因をユーザに提示することで空調機器の制御の根拠を明確化する。



4 Shapelets 学習によるインフラ・製造分野向け時系列波形の異常診断技術 — 異常の検知や診断に有効な波形パターンを発見する AI —

山口晃広 | (株) 東芝 研究開発センター

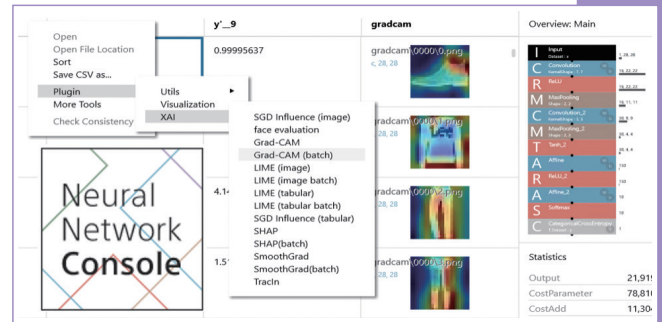
インフラ設備や製造装置のセンサから時系列波形データを収集し AI により異常や正常などの状態を自動判別する時系列波形異常診断技術に注目が集まっている。本稿では、インフラ・製造分野における技術課題として、AI の説明性、異常データの収集、誤判断のリスク管理、データの信頼性を述べる。次に、説明性のある波形診断技術として shapelets 学習法を説明し、残りの課題も解決するよう拡張した技術を産業分野への適用事例とともに紹介する。



5 説明可能な AI を身近にするためのディープラーニングツール

鈴木健二 | ソニーグループ (株) R&D センター

説明可能な AI を使いこなすためには、専門的な知識やプログラミング技術を必要とし敷居が高い。説明可能な AI とは、AI の判断根拠を人間が理解できるように可視化する技術である。本稿は、AI の判断根拠の可視化や認識精度の向上を GUI 環境にて手軽に操作できるディープラーニングツール「Neural Network Console」を紹介する。



6 信頼できる AI の実現に向けて — XAI による根拠の納得感向上のアプローチ —

恵木正史 間瀬正啓 濱本真生 | (株) 日立製作所 研究開発グループ

XAI (eXplainable AI) 技術の中でもデファクトの1つとなっている Shapley 値による説明方法について、その使いこなしのポイントを概観する。また、実案件への適用の過程で我々が遭遇した Shapley 法の2つの課題とその解決策について述べる。いずれも AI および XAI の信頼性にかかわる課題であり、現場から信頼される AI を実現する上で重要な論点になると考えられる。

