



特集 説明できるAI

AI判断の根拠を説明する日立の「XAI」 ～さまざまな業務領域で適用シーンを拡大～

人工知能 (AI) の利活用が広がるなか、「導き出された結論の根拠」が人には理解できず、重要な意思決定をともなう業務へのAI適用をためらうケースが増えています。そこで日立はAIを「ブラックボックス」化させないために、予測結果の根拠説明機能をアドオン (外付け) する新技術を開発。現場担当者やステークホルダーの納得感と信頼感を向上させ、幅広い業務領域でのAI活用に貢献します。

■ ブラックボックスといわれるAI

労働生産人口が減少し続ける日本では、慢性的な人手不足とともに、職場を離れるベテラン従業員による知識や技術の伝承も深刻な課題となっています。その解決策として、近年大いに注目されているのがAIです。

AIを活用することで、少ない人手でも非定型的な知的業務の作業効率が高まったり、品質維持に必要なスキルやノウハウを容易に継承したりすることができるからです。

実際にAIは、需要予測や品質管理、マーケティング、人材管理、リスク管理など、ビジネスの世界での利活用シーンを急速に広げています。

しかし、ディープラーニング (深層学習) は、学習結果に基づく「思考や判断」の過程が見えないため、AIの専門家でも導き出された結論の根拠が理解できず、AIが「ブラックボックス」であると認識される傾向があります。

このため、ステークホルダーへの説明が求められる、人命や財産に関わる重要な意思決定には「AIの適用が難しい」「リスクが大きい」といった不安の声が聞かれるようになっていました。

■ 判断根拠を明確にするXAI

例えば、金融分野における住宅ローン審査にAIを適用する場合、公正な判断であることの証明や、個別の審査結果の判断根拠をお客さまに説明することが要求されます。

また、病理診断のAIでは、判断根拠が明確でないと、病名の確定や患者との対話が深掘りできず、治療方針や投薬の判断を医師も患者も容易には受け入れられなくなります。

そこで、AIの判断根拠を人に説明できるようにすることで、人とAIの対話を促進し、AIへの信頼を高めていく技術 eXplainable AI (XAI) への期待が高まっています。

世界的にも研究が進んでいるXAIには、大きく2つのアプローチがあります。

1つは、学習内容や判断根拠がわかりにくい「ブラックボックス型AI」に、説明するための解釈機能をアドオンし、AIがなぜそのように予測したかを説明する技術です。もう1つは、最初からAIの学習過程や構造そのものを人が解釈しやすいようにつくる「トランスペアレント型AI^{※1}」です。どちらの技術についても、さまざまなベンダーや研究機関が積極的

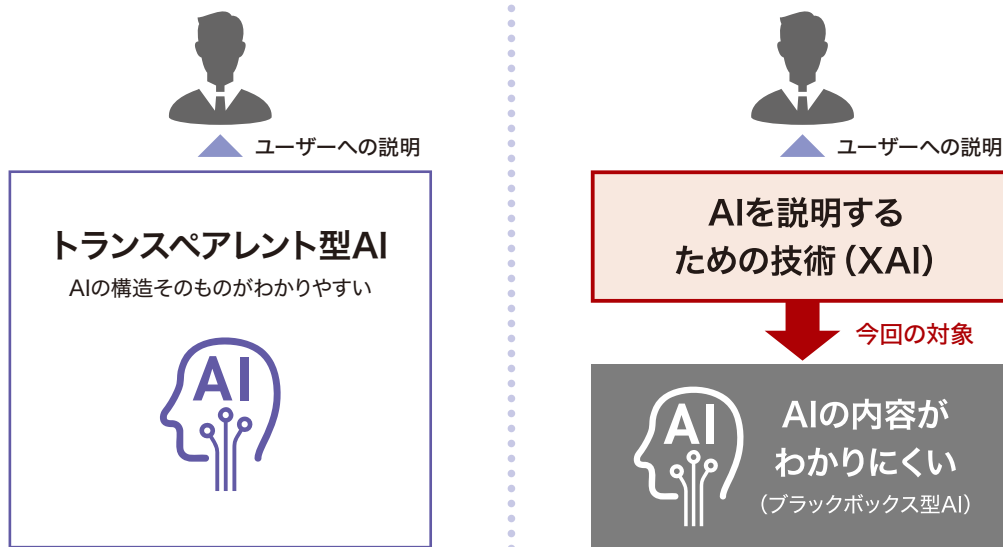


図1 XAIに対する2つのアプローチ

に開発を進めています。

※1 ブラックボックスに対して、中身が見え、透明性があるといわれるAI

■ ステークホルダーの納得感を向上させるXAI

日立は、多様な業態で事業を展開している実績と、そこで得たノウハウをベースに、さまざまな業務領域で多目的に使える人工知能「Hitachi AI Technology/H」を開発し、多くの案件に適用してきました。その過程で、金融与信審査や医療診断だけでなく、火災、交通事故、犯罪などの発生の予測で、その根拠をステークホルダーに提示する必要性と重要性の高まりを認識しました。またそのような現場で必要とされる根拠の観点も多様であることがわかってきました。その要望に応えるため、多様な観点からの根拠の提示を可能とするXAI技術を開発しました。

XAI技術は、さまざまなブラックボックス型AIに予測結果の根拠説明機能をアドオンするもので、お客さまニーズに合わせた多様な観点からの根拠説明の組み合わせにより、AIを活用した迅速な意思決定や知識の抽出、精度の改善、公平性の評価などを実現。わかりやすい根拠の提示により、現場担当者やステークホルダーの納得感と信頼感を向上させ、幅広い業務領域でのAI活用を加速します。

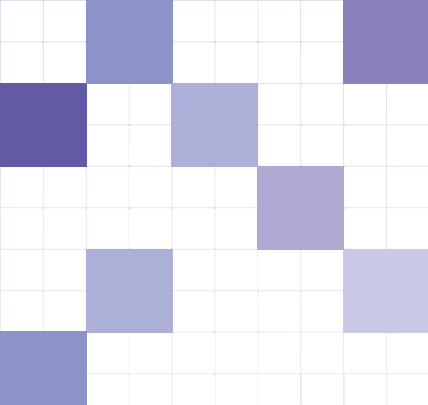
■ 金融機関との協創でローン審査に実運用

XAI技術の1つの応用として提供されるのが、「Hitachi AI Technology/Prediction of Rare Case(AT/PRC)」です。AT/PRCは、正常なデータの学習に加え、まれな事象の発生を予測することが特徴です。日立独自の「シグナル&ノイズに基づく学習」により、予測精度を下げる過学習を抑制し、予測結果の根拠を定量的に提示する「影響度算出技術」を用いています。貸し倒れ、株式の不正取引審査や新規取引顧客の評価など、金融機関をはじめとする企業が取り組むさまざまなリスク管理業務での適用にマッチしています。

このAT/PRCの適用例となるのが、2019年より実運用を開始した、住信SBIネット銀行 株式会社（以下、住信SBIネット銀行）と日立との協創による「AI審査サービス」です。

金融機関の与信審査は、お客さまの収入や資産の状況などを詳細に把握し、その情報をもとに将来を予測するための高度なノウハウが求められます。特に住宅ローンは取引期間が長期にわたるため、予測が非常に難しく、複雑なデータの相関を見いだす高度なアルゴリズムの構築が大きな課題となっていました。

そこで両社は2016年10月より、住宅ローンをはじめとするローンの審査業務にAIを活用する実証実験を開始しました。



サービス概要

- ① AI審査モデルでPD※を算出し金融機関に提供
- ② ローン商品の与信状況に関わるモニタリング業務
- ③ 与信戦略構築に係るコンサルティング など

※PD:Probability of Default 債務不履行の確率

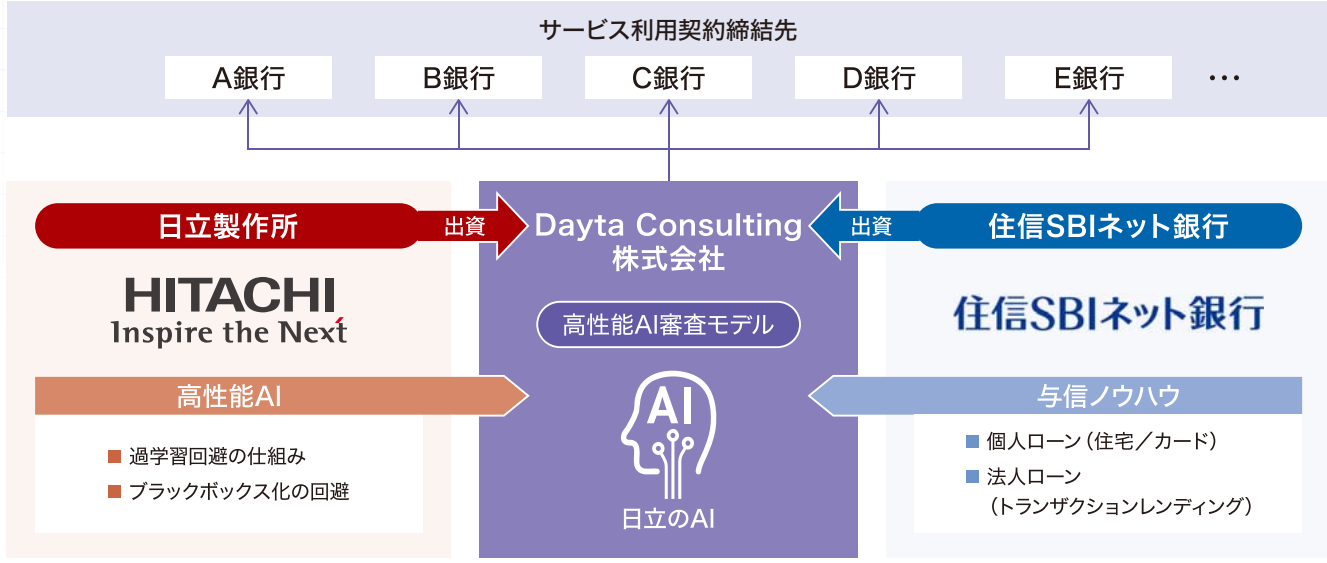


図2 住信SBIネット銀行と協創したAI審査サービス

AT/PRCのプロトタイプと、住信SBIネット銀行のデータハンドリング技術を組み合わせることで、従来のディープラーニングを用いたAIと比較して43%高い精度での予測を実現。精度向上※2と、住所、家族、年齢、年取といった属性ごとに影響度を分解して審査結果の根拠を説明できるサービスモデルを開発しました。

現在このAI審査サービスは、住信SBIネット銀行の住宅ローン審査業務で活用されているほか、同行と日立の合弁会社を通じて複数の金融機関にもサービスを提供する予定です。AI審査サービスは、大量データに基づく精緻でスピーディーな審査に加え、その根拠情報をわかりやすく提示することで、ベテラン従業員の不足や、業務の属人化に悩む多くの金融機関の課題を解消。より多くのお客さまの融資の要望に応じるだけでなく、信用コスト※3の削減にも貢献します。

※2 融資データを活用した貸し倒れ予測に適用した場合
 ※3 ある貸し出しについて、一定の期間中に回収不能になる可能性を考慮した平均的な損失見込み額

消防機関と「救急需要予測システム」を協創

このほかにも日立は、XAI技術が生成する判断根拠の妥当性や有用性を、さまざまな有識者や現場担当者の検証を経たうえで、犯罪・事故予測、火災リスク予測といった公共分野の実案件に適用し、AI活用の可能性を着実に広げています。

その一例が、日本の消防機関と協創した「救急需要予測システム」です。高齢化が進む日本では、救急出動件数が過去20年間で約80%増加し、救急リソースの不足により現場到着時間が約40%（約2分半）伸びています。現場到着時間の延伸は救命率の低下に直結するため、各消防機関は現場到着時間の短縮きっさんを喫緊の課題としています。

一部の消防機関では、日中はオフィス街に、夜間は繁華街に救急隊を待機させるといった機動的な運用を開始しており、現場到着時間の短縮に一定の効果を上げています。

こうした運用をさらに効果的かつ広範に行うためには、データに基づいた予測によって現場職員を支援することが不可欠です。日立と消防機関は、天候や人口分布などを

とに、1kmメッシュごとの事案発生件数と現場到着時間を予測、管轄地区全体の平均現場到着時間が最短になる救急隊員の配置をAIがRecommendするシステムを開発。XAIの技術を使い、地図上のメッシュをクリックすると、該当箇所の予測において、AIがどのようなデータをどの程度重視したかをわかりやすく表示する機能を設けました。

これにより現場職員がAIの予測への理解を深め、信頼できるようになり、安心感を持って行動に移し、本来業務である救急患者の救命に注力できるようになりました。

今後も日立は、AI判断の根拠を説明するXAI技術の精度と機能を継続的に強化しながら、お客さまのビジネスを成功に導き、社会課題のよりよい解決を支援するAI活用の可能性を広げていきます。

次ページでは、日立のXAI技術を活用し、お客さまの継続的なAI運用と改善を支援する「AI導入・運用支援サービス」を紹介します。

**日立は、金融・公共など多数の実案件に適用、
XAIが生成する判断根拠の
妥当性や有用性を有識者が検証**



与信審査



犯罪・事故予測



救急需要予測



火災リスク予測

図3 お客さまとの協創によるビジネス適用・有効性検証

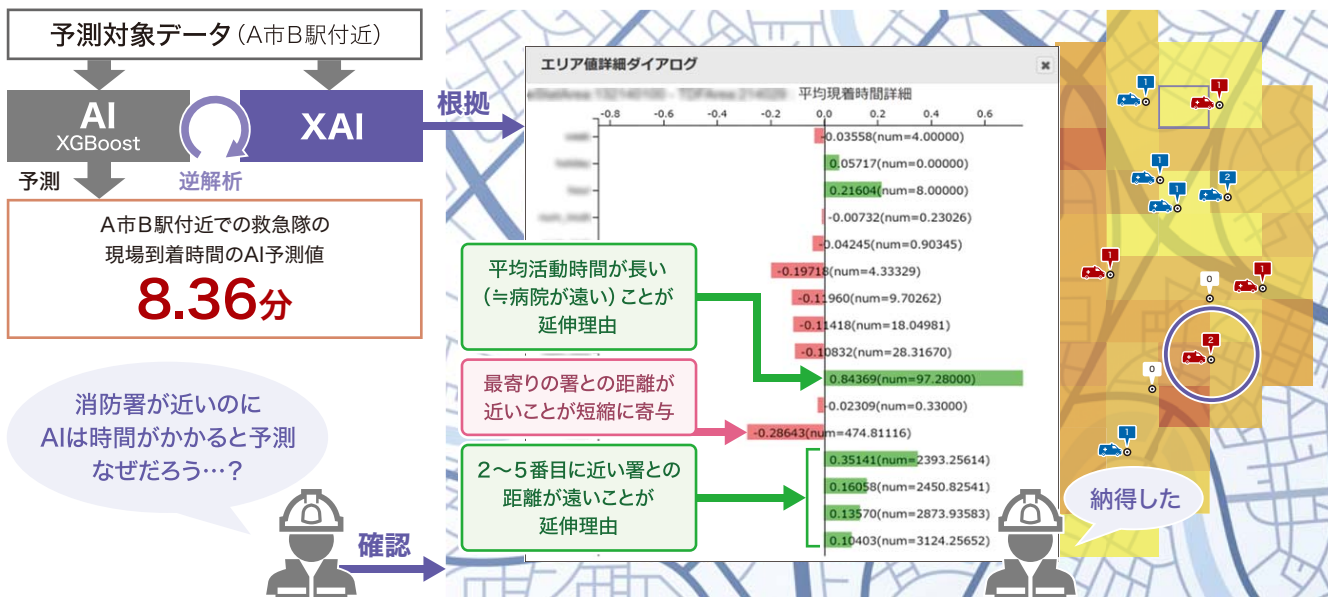


図4 救急需要予測AIシステムの概要

お問い合わせ先・情報提供サイト

(株)日立製作所 研究開発グループ デジタルテクノロジーイノベーションセンター データサイエンス研究部
<https://www.hitachi.co.jp/rd>

