



望む結果までの手順を導くことができる「説明可能な AI」を 世界で初めて開発

人と協働する AI の信頼性と透明性の向上を実現

株式会社富士通研究所(注1)(以下、富士通研究所)と国立大学法人北海道大学(注1)(以下、北海道大学)は、AIが自動判断した結果を基に、望む結果を得るために必要な手順を自動で提示できる技術の世界で初めて開発しました。

AI分野では、データから自動的に判断するだけでなく、個々の判断理由を与えてくれる「説明可能なAI」や、望ましくない判断結果に対して、改善するための項目を提示するAI技術が注目されています。しかし、これらの技術は、個々の項目に関して、これをしてあげれば良い結果が得られたという仮定の改善項目は示してくれますが、判断結果を改善していくための手順などは提示してくれませんでした。

例えば、健康状態を判断するAIが不健康と判断した場合、身長や体重、血圧などの健康診断項目のデータから不健康の判断理由を説明するだけでなく、健康になるための最善策を提示してほしいというニーズに対して、本技術を適用することで、過去データから複雑に絡み合った多数の診断項目間の相互作用を特定し、実現の可能性や実施の難易度を考慮した上で具体的な改善手順を示すことができます。

これにより、AIが出した判断理由を知るだけでなく、個々の利用者が望む結果を得るために取るべき改善の手順を示すことが可能となり、AIによる判断の信頼性と透明性が向上するとともに、今後、人と共に働くAIの活用の幅を広げることが期待できます。

本技術の詳細については、2月2日(火曜日)から開催されているAI分野の難関国際会議「AAAI-21 (The Thirty-Fifth AAAI Conference on Artificial Intelligence)」にて発表します。

【 開発の背景 】

現在、顔認証や自動運転など高度なタスクが求められるAIシステムに広く用いられている深層学習技術は、大量のデータに基づいた様々な判断を予測モデルと呼ばれる一種のブラックボックス的な規則を用いて自動的に行っています。しかしながら、今後、社会の重要な判断や提案をAIが担うためには、AIシステムの透明性と信頼性の担保が課題となってきました。このような社会と技術の動向から、データに基づいて自動的に判断するだけでなく、個々の判断理由を提示する「説明可能なAI」と呼ばれる新しいAI技術の研究が盛んになっています。

例えば、健康診断などにおいて、体重や筋肉量などのデータを基に病気になるリスクの高低を、AIによって判断することが可能になっていますが(図1(A))、リスクの高低に関する判断結果に加えて判断の根拠となった属性(図1(B))を提示する「説明可能なAI」が注目されています。

AIは、入力データの属性を基に健康リスクが高いと判定しているため、健康リスクが低いという望む判断結果を得るためには、これらの属性の値を変えればよいと言えます。

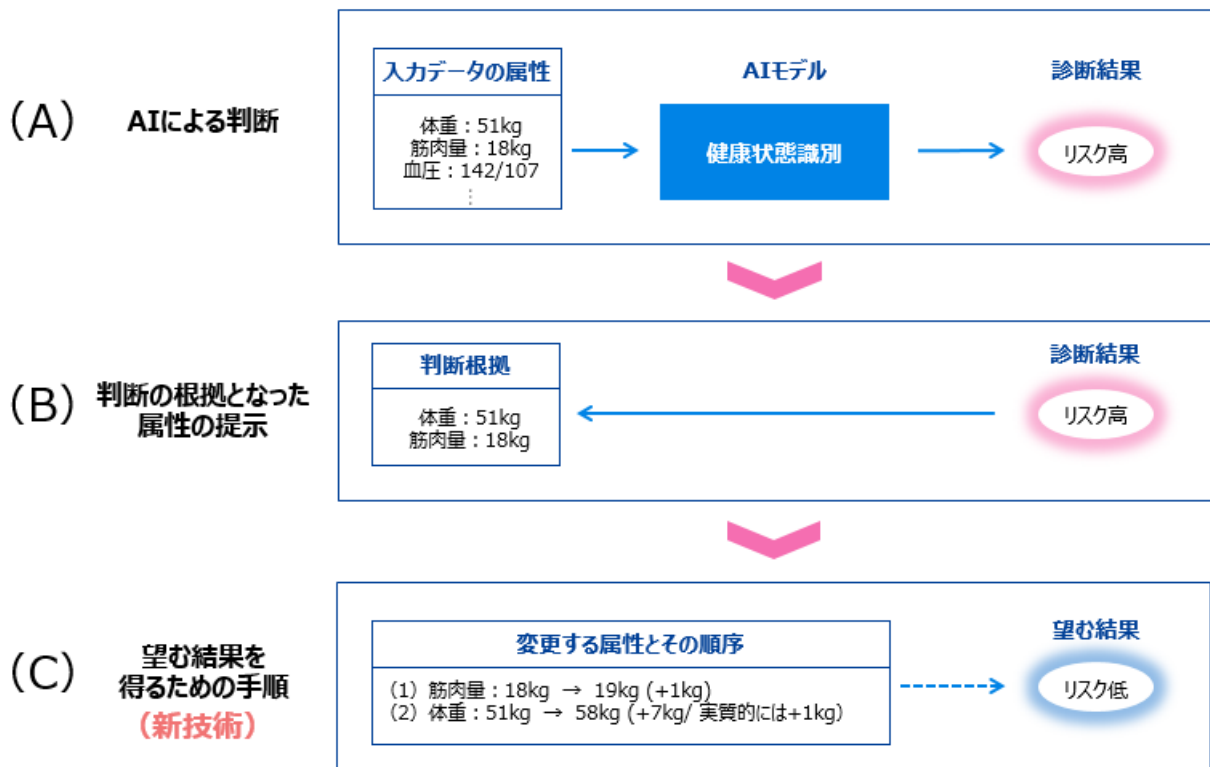


図 1 AI の活用方法と説明機能

【 課題 】

AIによる自動判断において、望む結果を得るためには、変更が必要な属性を提示するだけでなく、その変更が現実的かつできるだけ小さい労力で行える属性を提示することが必要になります。

健康診断の例においては、AIによる判断結果をリスク高の現状からリスク低に変えたい場合、筋肉量を増やせば少ない労力で行えそうですが(図2 変更1)、体重を変えずに筋肉量だけを増やすことは非現実的であるため、実際には体重と筋肉量を増やすことが現実的な解(図2 変更2)であると言えます。また、体重や筋肉量などの属性の間には、筋肉を増やすと体重も重くなるというような因果関係などの相互作用も多く、変更にかかる総労力は、各属性の変更順序に依存するため、属性の適切な変更順序を提示する必要があります。図2では、現状から変更2に到達するために、体重と筋肉量のどちらを先に変えればよいかは自明ではないため、膨大な変更点の候補の中から実現の可能性や順序を考慮した上で、適切な変更の方法を見つけることが課題となっています。

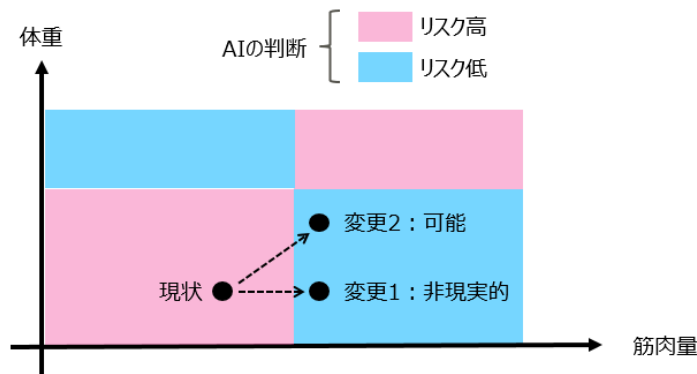


図2 属性の変更

【 開発した技術 】

これまで富士通研究所と、北海道大学大学院情報科学研究院の有村研究室は、機械学習とデータマイニングに関する共同研究により、AIの判断理由を人間に説明できる新しいAI技術を開発し、人間にとって役立つ知識の発見に繋げてきました。

人の意思決定を支援するためのAI技術としてこれまでに開発されてきたLIME(注3)(注1)やSHAP(注4)といったAI技術は、AIがなぜこのように判断したかを説明することで、その判断結果に納得性を与える技術でした。今回、両者による共同研究において開発した新技術は、反実仮想説明(注5)という考えに基づき、属性変更におけるアクションとその実施順序を手順として提示します。過去の事例の分析を通して非現実的な変更を避けつつ、属性値の変更がほかの属性値に与える因果関係などの影響をAIが推定し、それに基づいて実際に利用者が変更しなければならない量を計算することで、適切な順序、かつ一番少ない労力で最適な結果が得られるアクションの提示を可能としました。

例えば、健康診断で望む結果にするために変更する入力属性とその順序(図1(C))において、リスクを低くするためには、筋肉量をプラス1kg、体重をプラス7kg変更しなければならない場合に、筋肉量と体重の間の相互作用を事前に分析することにより、筋肉量を1kgプラスすれば体重は6kgプラスされるというような関係が因果関係の分析により推定できます(図3)。その場合、体重の変化量として必要なプラス7kgのうち、筋肉量の変更の後に必要となる変化量はプラス1kgとなります。つまり、実際に変化させなければならない変更量は、筋肉量プラス1kgと体重プラス1kgであると言えるため、先に体重を変化させるよりも少ない労力で望む結果を得ることができます。

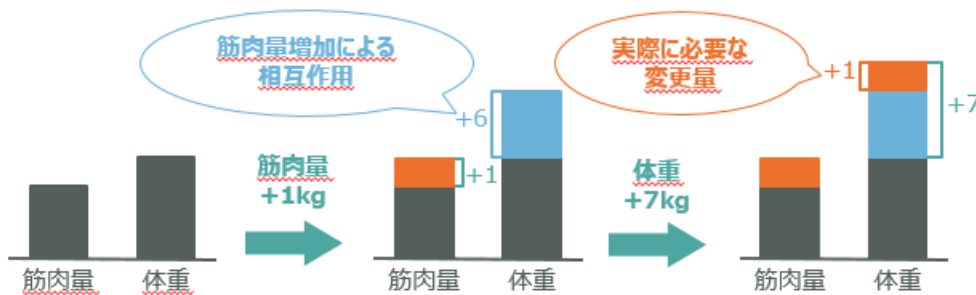


図 3 属性間の相互作用と変化量

【 効果 】

今回、共同開発した反実仮想説明AI技術を用いて、この分野で主に用いられる糖尿病、ローンの与信審査、ワインの評価の3種類のデータセットにて検証(注6)を行いました。機械学習の主要なアルゴリズムであるロジスティック回帰(注7)、ランダムフォレスト(注8)、多層パーセプトロン(注9)の3つのアルゴリズムと開発技術を組み合わせ、AIの判断結果が望ましくない場合に、望む結果を得るためのアクションの提示を目的とした検証を行った結果、今回の開発技術が、すべてのデータセットおよび機械学習アルゴリズムの組み合わせにおいて、少ない労力で推定結果を望む結果に変更するための適切なアクションと実施順序を取得できたことを確認し、特にローンの与信審査のケースでは、半分以下の労力を実現しました。

本技術を活用することで、AIによる自動判断において望ましくない結果が予想された場合に、その結果を望む結果に変えるために必要なアクションを提示することが可能となります。これにより、AIの用途を判断だけでなく人の意思行動の支援に拡げることができ、AIの適用先のさらなる拡大が期待できます。

【 今後 】

今後、富士通研究所は、個別の因果関係を発見する技術(注10)と組み合わせることで、より適切なアクションを提示できるように継続して取り組みます。そして独自開発した「FUJITSU AI Technology Wide Learning」によるアクション抽出技術(注11)を本技術により拡張し、富士通株式会社(注12)のAI技術「FUJITSU Human Centric AI Zinrai」を支える新たな機械学習技術として2021年度の実用化を目指します。

北海道大学は、アクションの提示に限らず、多様な現場のデータから人間の意思決定に役立つ知識や情報を抽出するためのAI技術の確立を目指します。

【 商標について 】

記載されている製品名などの固有名詞は、各社の商標または登録商標です。

以 上

【 注釈 】

- (注1) 株式会社富士通研究所:本社 神奈川県川崎市、代表取締役社長 原 裕貴。
- (注2) 国立大学法人北海道大学:所在地 北海道札幌市、総長 寶金清博。
- (注3) LIME: AI の説明技術の一つ。解釈可能なシンプルなモデルによって説明する。
- (注4) SHAP: AI の説明技術の一つ。モデルにおける説明変数の寄与度を示すことにより説明する。
- (注5) 反実仮想説明:「もしこれをしていれば結果はこうなっていた」というように、事実とは異なる状態を示し説明する手法。
- (注6) データセットにて検証: 機械学習の比較評価用に数々のデータセットを提供している世界的に有名なりポジトリである UC Irvine Machine Learning Repository、およびクレジットスコアリング会社である FICO 社が機械学習のために公開したデータのうち、糖尿病、ローンの与信審査、ワインの評価の3種類のデータセットにて検証。
- (注7) ロジスティック回帰: 機械学習アルゴリズムの一種。超平面にロジスティック関数を組み合わせた確率モデル。
- (注8) ランダムフォレスト: 機械学習アルゴリズムの一種。多数の決定木分類器で多数決を行って安定した判断をする予測モデル。
- (注9) 多層パーセプトロン: 機械学習アルゴリズムの一種。多層のニューラルネットワークを学習するモデル。
- (注10) 個別の因果関係を発見する技術: ヒトやモノなどのデータの一つひとつが持つ特徴的な因果関係を発見する技術を開発(2020年12月17日プレスリリース) (<https://pr.fujitsu.com/jp/news/2020/12/17.html>)
- (注11) 「FUJITSU AI Technology Wide Learning」によるアクション抽出技術: 最適なアクションプランを提案、「Wide Learning」の新技術を開発(2019年9月13日プレスリリース) (<https://pr.fujitsu.com/jp/news/2019/09/13.html>)
- (注12) 富士通株式会社:本社 東京都港区、代表取締役社長 時田隆仁。

【 関連リンク 】

- ・ 富士通研究所 技術トピックス「反実仮想説明に基づくアクションの最適順序を推薦する新たなAI技術を開発」(<https://www.fujitsu.com/jp/group/labs/about/resources/article/202102-aaai-21.html>)
- ・ Wide Learning公式サイト「Hello, Wide Learning!」(<https://widelearning.labs.fujitsu.com/>)

《 本件に関するお問い合わせ 》

株式会社富士通研究所

人工知能研究所

電話: 044-754-2328 (直通)

E-mail: widelearning@ml.labs.fujitsu.com

《報道関係者お問い合わせ》

富士通株式会社

広報IR室

電話:03-6252-2174(直通)

北海道大学

総務企画部広報課

電話:011-706-2610

E-mail: kouhou@jimuhokudai.ac.jp