

PRESS RELEASE (2026/05/11)

**AIが「なぜ異常と判断したか」を波形で示す反事実波形生成技術を開発
～インフラ設備や製造装置の異常診断で、
原因究明と保守判断を支える説明性の高いAIの実現に向けて～**

九州大学マス・フォア・インダストリ研究所（以下、IMI）は、株式会社東芝との共同研究により、インフラ設備や製造装置の異常検知AIにおいて、「AIがなぜ異常と判断したのか」を波形の違いとして可視化できる「反事実波形生成技術」を開発しました^(*)。本技術は、AIが異常と判定したセンサ波形に対し、正常と判定されるためにはどの部分をどのように変えればよいかを波形として生成するものです。これにより、AIの判定根拠を局所的な波形の違いとして可視化でき、現場の専門家が判定理由を直感的に理解できます。本技術は、説明可能AI（XAI）における反事実説明の考え方を時系列波形データに応用したものです。従来、時系列波形データを対象とする多くの説明可能AIはニューラルネットワークを前提としていましたが、本技術は、ニューラルネットワークよりも高性能な判定AIや、異常データを事前に収集しにくい異常検知にも適用できます。公開時系列データセットで評価した結果、同種の従来技術と比べて、元波形からの不要な変形を抑えた高品質な反事実波形を生成できることを確認しました。

本技術の詳細については、5月2日から5日にかけてタンジェ（モロッコ）で開催される機械学習・統計科学分野におけるトップクラスの国際会議であるAISTATS（International Conference on Artificial Intelligence and Statistics）において、5月3日に発表しまし

【研究の背景と経緯】

近年、インダストリアルIoTの普及に伴い、設備や装置に取り付けたセンサから時系列データを収集し、AIで異常を高精度かつ早期に検知する取り組みが広がっています。特にインフラ・製造分野では、高い安全性や信頼性が求められるため、判定精度だけでなく、専門家が判定根拠を理解できることが重要です。さらに、現場の専門家は、波形を見て特徴量を作成し異常検知を行ってきた経験が過去にあるため、波形に対して専門知識があり、「なぜ／どのようにAIが異常と判断したか」を波形から解釈することを求めています。

AIによって時系列データの正常・異常を判定する技術はベンチマークが盛んでよく研究されており、ニューラルネットワークよりも高性能な手法が多数知られています。一方で、従来の説明可能AI（XAI）の多くは、画像や表データを対象としたニューラルネットワークへの適用を前提としており、時系列データを対象とするニューラルネットワークよりも高性能なモデルには適用しにくいという課題がありました。また、信頼性の高い製造装置やインフラ設備では異常データを事前収集しにくいという課題もあるため、正常データだけで学習する「異常検知」にも適用できる説明可能AI（XAI）が求められていました。

【研究の内容と成果】

本技術では、説明可能AI（XAI）の反事実説明の考え方を時系列データへ拡張し、AIで異常と判定された波形に対して、正常判定に切り替わるために必要な波形変化を生成します。反事実説明とは、入力データのどの部分をどのように変えれば、AIの判定結果が変わるかを示す考え方です。本技術に

より、AI の判定に影響を与えた波形の主要な箇所を特定できるだけでなく、正常と異常の境界で波形がどのように崩れるのかを、「反事実波形」と呼ばれる判定根拠として具体的に示すことができます。

本技術は、まず学習段階と運用段階で役割を分けます（図 1）。学習段階では、収集したデータは全て正常であるため、深層学習を用いた 1 クラス分類（※1）についての既存研究のアイデアにならない、データから抽出した特徴量を正常として、凝集したクラスタを形成するように学習します。運用段階では、異常と判定されたデータの特徴量を正常と学習したクラスタの重心に近づけることで、どの方向にどれだけ動けば、正常波形に対応するか分かります。また、学習段階では時間順序構造を維持するように学習します。これにより、運用段階では、異常と判定されたデータの特徴量を正常と学習したクラスタの重心に近づけて変更箇所を減らすことが可能となり、元波形から不要な箇所の変化を抑えられます（図 2）。これらは離散最適化問題に定式化でき、この問題に対する高速な解法も開発しました。これにより、AI の異常判定の根拠を局所的な波形の崩れとして提示する、説明性の高い反事実波形を高速に生成できるようになりました。本成果は、2024 年 10 月から開始した株式会社東芝との共同研究によるものです。

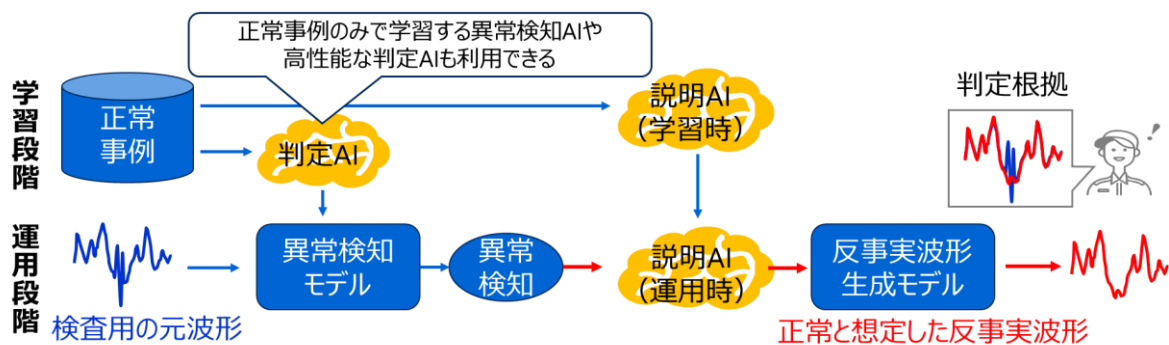


図 1：反事実波形生成技術の概要

ベンチマーク評価では、高性能な判定 AI である QUANT（※2）を用いて、異常検知およびクラス分類の両設定で 26 種類の時系列データセットを評価しました。その結果、従来技術と比べて、元波形からの不要な変形が統計的に有意に抑えられることが確認され、高品質な反事実波形を生成できることがわかりました。

さらに、本技術は判定 AI と説明 AI を分離しているため、2018 年に株式会社東芝が発表した「正常時の波形データのみで学習し異常を検知する説明性の高い AI」である OCLTS（※3）と比べて、最新の高性能な異常検知モデルを組み合わせやすいという特長があります。加えて、異常と判定した時系列データを説明 AI の学習に利用できるため、より高品質な説明生成も期待できます。

今回の開発についてのポイント

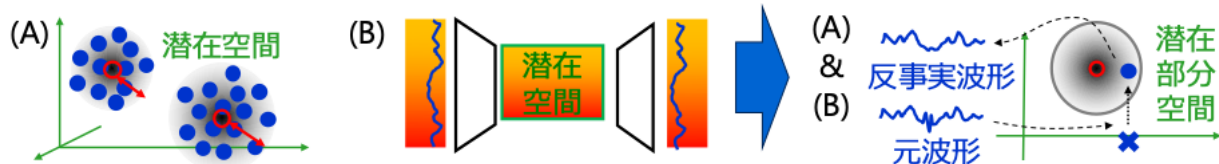
正常事例だけで潜在構造を学習しておき 運用時に高品質な反事実波形を効率的に生成

学習段階

- A) 正常クラスの潜在表現がクラスタを形成するように潜在空間を学習
潜在空間にコンパクトな正常領域を学習する深層1クラス分類に相当
- B) 時間順序構造を維持して潜在部分空間を学習

運用段階

- A) 勾配を用いずに クラスタの重心等に向かって 反事実波形の潜在表現を効率的に特定
- B) 潜在部分空間での変更箇所を減らすことで 反事実波形を元波形から局所的に変更



© 2026 Toshiba Corporation 1

図2：本研究における技術（学習段階・運用段階の概要）

【今後の展開】

九州大学 IMI は、株式会社東芝による「安全性や信頼性が求められるインフラ設備や製造装置の異常検知に対する本技術の適用の推進、早期の実用化」を促進・発展させるための更なる共同研究開発を実践します。

【参考資料】

(*1) 九州大学プレスリリース

<https://www.kyushu-u.ac.jp/ja/notices/view/2860>

株式会社東芝プレスリリース

<https://www.global.toshiba/jp/technology/corporate/rdc/rd/topics/25/2503-02.html>

【用語解説】

(※1) クラス分類

正常なデータだけを使って AI で基準を学習し、その基準から外れるデータを異常として判定する方法。

(※2) QUANT

生の時系列データに加えて 1 階差分、2 階差分、フーリエ係数といった複数の系列表現を用いて、それぞれを複数の固定区間に分割した上で、各区間における分位点を特徴量として抽出し、Extra Trees 分類器により分類を行う時系列分類手法。Dempster et al. (2024)

<https://doi.org/10.1007/s10618-024-01036-9> で提案された手法。

(※3) One-Class Learning Time-series Shapelets (OCLTS) :1 クラス分類向け時系列シェイプレット

詳細は以下に記載：正常時の波形データのみで学習し異常を検知する説明性の高い AI を開発

<https://www.global.toshiba/jp/news/corporate/2018/12/pr1301.html>

株式会社東芝プレスリリース 2018 年 12 月 13 日

【謝辞】

本研究は九州大学 IMI と、株式会社東芝の共同研究に基づくものです。

【論文情報】

国際会議名 : The 29th International Conference on Artificial Intelligence and Statistics (AISTATS 2026)

タイトル : Counterfactual Explanations via Latent Structure for Time Series Classification

著者名 : Akihiro Yamaguchi, Shizuo Kaji, Kaname Matsue, Ryusei Shingaki

Kyushu
University

VISION 2030

総合知で社会変革を牽引する大学へ

【お問い合わせ】九州大学 マス・フォア・インダストリ研究所 教授 松江 要
TEL : 092-802-4420
FAX : 092-802-4405 (数理・MI 研究所事務室)
Mail : kmatsue@imi.kyushu-u.ac.jp