



Title: CausalAIME: Leveraging Peter-Clark Algorithms and Inverse Modeling for Unified Global Feature Explanation in Healthcare

(CausalAIME：医療分野における統一的なグローバル特徴説明のためのピーター・クラークアルゴリズムと逆モデルリングの活用)

Authors: Takafumi Nakanishi

(中西崇文(東京工科大 コンピュータサイエンス学部 教授)

Journal: R. Guidotti et al. (Eds.): xAI 2025, CCIS 2577, pp. 332–356 7

掲載年月: 2026年10月

研究概要：医療分野での機械学習応用において、モデルの挙動解釈と特徴量と臨床結果の真の関連性の探索が重要です。従来の説明可能 AI (XAI) は「モデル依存」の大域的特徴量重要度に焦点を当てており、因果関係や医学的整合性を保証しませんでした。そこで本研究では、近似逆モデル説明 (AIME) と因果探索のための Peter-Clark (PC) アルゴリズムを組み合わせた新フレームワーク「CausalAIME」を提案しました。これにより多重共線性を抑制しつつ、特徴量の符号 (正・負) とクラス別の寄与を可視化することに成功しました。さらに、モデルの出力または真のラベルを入力として選択することで、モデル依存とデータ駆動の両方の重要度を統一的に扱うスキームを実現しました。乳がん診断データを用いた実験により、CausalAIME が臨床現場で重要な符号ベースの解釈性と因果的視点を提供することを示しました。

研究背景：近年、医療分野で機械学習の導入が進んでいますが、高精度なモデル構築だけでなく、予測理由の説明が信頼確立のために不可欠です。SHAP や Random Forest などの既存手法は、学習済みモデルの挙動に依存する「モデル依存」の重要度を提供しますが、真のラベルとの因果的・統計的関係は保証されません。医療現場では、モデルバイアスを含まない「データ駆動（モデル非依存）」の重要度が求められることが多くあります。既存の AIME はこれを扱えますが、特徴量間の因果構造や多重共線性の考慮が不十分でした。そこで、AIME に PC アルゴリズムを統合し、特徴量間の因果・相関関係を考慮した大域的特徴量重要度を導出する必要がありました。

研究成果：乳がん診断データセット (Breast Cancer Wisconsin) を用いた比較実験により、以下の成果を得ました 17。医学的整合性と多重共線性の抑制：PC アルゴリズムにより推定された因果・相関構造に基づくペナルティ項を導入することで、AIME 単体では多重共線性により生じていた符号の反転 (直感に反する寄与) が抑制されました。その結果、腫瘍の大きさや形状の不規則性が悪性度と正の相関を示すという、既存の医学的研究と整合する結果が得られました。定量的性能の優位性：記述精度 (Descriptive

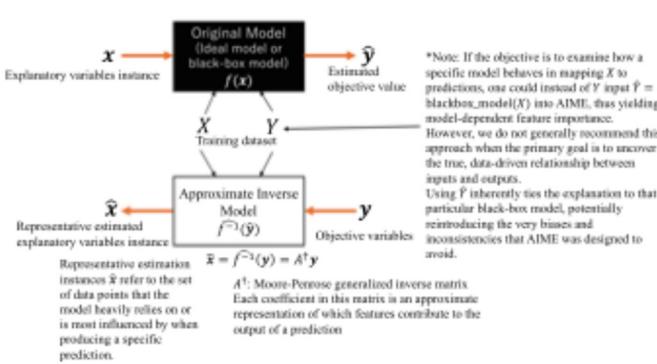


図 1 AIME の概要。学習済みモデル（または理想モデル）の入出力関係の逆写像を近似することで、特徴量の寄与を推定するフレームワーク。モデル依存（予測値 \$Y\hat{Y}\$ を使用）とデータ駆動（真値 \$Y\$ を使用）

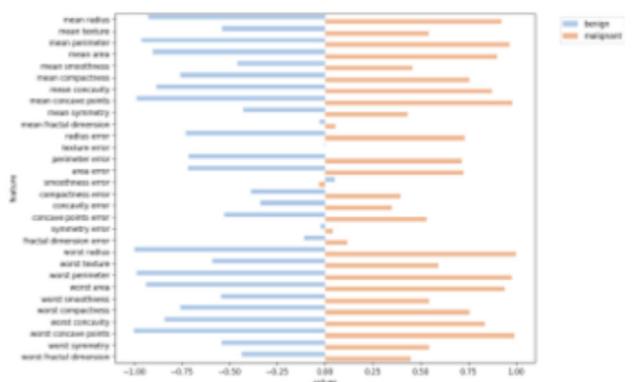


図 2 CausalAIMEによる大域的特徴量重要度の結果（乳がんデータセット）。PC アルゴリズムによるペナルティ項の導入により、多重共線性が抑制され、医学的知見と整合する結果（例：腫瘍の大きさや形状の不規則性が悪性度と正の相関を示す）が得られている。

Accuracy)、安定性、効率性など6つの指標を用いた評価において、CausalAIMEは記述精度（0.88）でAIME（0.91）に次ぐ高い値を達成し、SHAP（0.75）を大きく上回りました。また計算時間においてもSHAPより高速（7.6秒 vs 42.2秒）であり、精度と計算コストのバランスが取れた手法であることが実証されました 21212121。データ駆動とモデル依存の統合：真のラベル（Y）を用いた「データ駆動モード」により、モデルのバイアスに左右されない真の関連性を抽出できる一方で、必要に応じてモデル出力（ \hat{Y} ）を用いた「モデル依存モード」へ切り替え可能な柔軟性を持つことが確認されました。

社会への影響：医療などの信頼性が重視される分野において、AIモデルの挙動解釈だけでなく、データ背後にある真の関連性（因果関係）を理解することが可能となります。これにより、医師や専門家が納得できる医学的整合性のある根拠に基づいた意思決定支援が可能となり、説明可能性の向上に貢献します。

専門用語：

CausalAIME：本研究で提案する手法。AIMEの逆モデル近似に、PCアルゴリズムで推定した特徴量間の因果・相関構造をペナルティ項として組み込んだ統合フレームワーク。

Peter-Clark (PC) アルゴリズム：条件付き独立性検定を用いて、変数間の因果グラフ（DAG）のスケルトンと枝の向きを探索する代表的な因果探索アルゴリズム。

近似逆モデル説明 (AIME)：ブラックボックスモデルの出力から入力を推定する逆写像（逆演算子）を数学的に近似することで、大域的特徴量重要度を算出する手法。

モデル依存 / データ駆動：「モデル依存」は特定の学習済みモデルの挙動に基づく重要度。「データ駆動」はモデルのバイアスを除いた、入力データと真のラベルとの純粋な関係に基づく重要度。