

交互作用を考慮した SHAP による 機械学習モデルの解釈

野原 康伸^{*1}, 井口 登興志^{*2}, 野尻 千夏^{*3}, 中島 直樹^{*3}

^{*1} 熊本大学大学院, ^{*2} 福岡市健康づくりサポートセンター, ^{*3} 九州大学病院

Explanation of Machine Learning Models Using SHAP Considering Interaction Effects

Yasunobu Nohara^{*1}, Toyoshi Inoguchi^{*2}, Chinatsu Nojiri^{*3}, Naoki Nakashima^{*3}

^{*1} Kumamoto University, ^{*2} Fukuoka Health Promotion Support Center,

^{*3} Kyushu University Hospital

抄録: 近年、機械学習手法の医療分野への応用が進んでおり、なぜそのような結果が出たのかを解釈・可視化する SHAP が注目を集めている。複数の変数の組み合わせによる交互作用は、アウトカムに大きな影響を与えることがあり、機械学習モデルを理解するのに重要であるが、これまでの SHAP 解析では、説明変数単体がアウトカムに与える影響と、説明変数同士の交互作用をまとめて評価していた。本論文では SHAP 値の導出アルゴリズムを改良し、説明変数単体がアウトカムに与える影響と説明変数同士の交互作用を分離して機械学習モデルを解釈する手法を提案する。改良手法を九州大学病院の癌コホートデータ(N=29080)に適用し、どのような要因の組み合わせが癌発生リスクに寄与しているのかを解析した。その結果、ビリルビン高値による大腸癌発症リスクの減少は、高齢者ほど特に効果が大きいことが分かった。

キーワード SHAP 機械学習 解釈 交互作用 Cox 比例ハザードモデル

1. はじめに

近年、深層学習をはじめとする機械学習手法の医療分野への応用が進んでいるが、機械学習モデルでは中身がブラックボックスの場合が多く、なぜそのような結果が出たのかの説明できない点が大きな課題とされてきた。SHAP (SHapley Additive exPlanation)[1,2]は、機械学習の結果を解釈する手法の一つであり、機械学習の出力を各説明変数の貢献度に応じて決まる SHAP 値の線形和の形で表現する。SHAP により、説明変数とアウトカムの間の非線形な関係を図示できるだけでなく、SHAP 値の傾き(微分)が、一般化線形モデルの偏回帰係数に対応するなど、既存の解釈方法との親和性も高い。

交互作用とは、ある説明変数が目的変数に与える影響が、他の説明変数の値によって異なることである。単独の説明変数だけでは、目的変数にあまり影響を及ぼさない場合であっても、その交互作用によっては目的変数に対して思わぬ大きな影響を及ぼすことがある。交互作用を理解することは、機械学習モデルを理解するのに重要であるが、これまでの SHAP 解析では、説明変数単体がアウトカムに与える影響と、説明変数同士の交

互作用をまとめて評価していた。このため、どの説明変数が重要かは分かっても、変数単体がアウトカムに影響を与えているのか、交互作用が影響を与えているのか分からない問題があった。

本論文では SHAP 値の導出アルゴリズムを改良し、説明変数単体がアウトカムに与える影響と説明変数同士の交互作用を分離して機械学習モデルを解釈する手法を提案する。改良手法を九州大学病院の癌コホートデータ(N=29080)に適用し、どのような要因の組み合わせが癌発生リスクに寄与しているのかを解析した。

2. 方法

これまでの SHAP 値の導出アルゴリズムを改良し、Shapley-Taylor Index[3]を用いることで、説明変数単体がアウトカムに与える影響と説明変数同士の交互作用を分離して導出する。改良手法を九州大学病院の癌コホートデータ(N=29080)[4]に適用し、どのような要因の組み合わせが癌発生リスクに寄与しているのかを解析した。

2.1 データについて

九大病院に入院したことのある 20 歳から 69 歳の患者で適用基準を満たす者の 10 年間の医療記録を電子カルテから取得した(N=29,080) [4]。

2.2 分析手法について

7つのベースライン特性を説明変数、大腸癌の発症を目的変数として、Cox 比例ハザードモデルによる生存時間分析を実施した。一般に Cox 比例ハザードモデルでは、ベースラインハザード関数としてロジスティック回帰を適用するが、本解析では機械学習モデルの代表的手法である Gradient Boosting Decision Tree(GBDT)を適用した。作成した GBDT によるベースラインハザード関数を SHAP により分析し、どのベースライン特性の組み合わせが癌発生リスクに寄与しているか調べる。

3. 結果

GBDT により作成した大腸癌発症予測モデルの時間依存型 ROC を描き、その AUC の 10 年間平均を調べると 0.662 であった。Shapley-Taylor Index により、各説明変数の単独効果と交互作用項に分離した場合において、どの項目の重要度が高いかを SHAP 値の標準偏差で評価した。単独作用項のみに注目すると、重要度の順位が高い順に、年齢、ビリルビン、中性脂肪、BMI、喫煙歴有であり、既存手法の順位と一致していた。

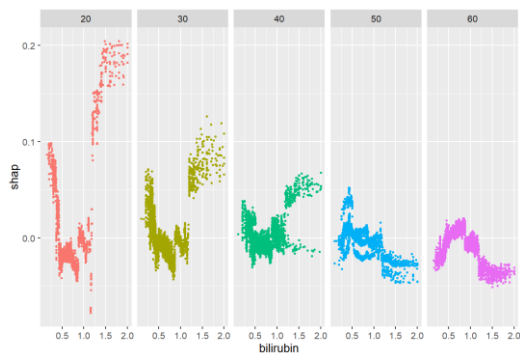


図 1 ビリルビンと年代の交互作用の SHAP Plot

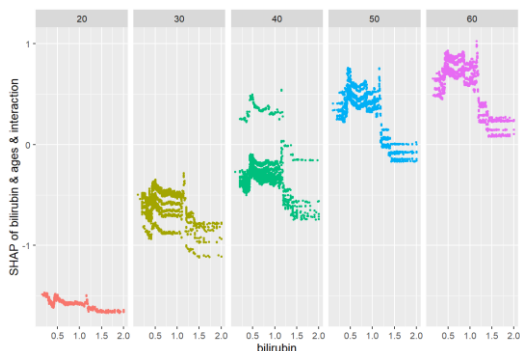


図 2 ビリルビンと年代およびその交互作用の和の SHAP Dependence Plot

4 考察

図 1 にビリルビンと年代の交互作用(重要度 9 位)に関する SHAP Dependence Plot を示す。若い年代であるほどビリルビンが 1.2mg/dl を超えると、交互作用に関する SHAP 値が大きく増加しているのに対して、50 代と 60 代では 1.2mg/dl を超えると SHAP 値が若干減少している事が分かる。

図 2 にビリルビン(重要度 2 位)と年齢(同 1 位)の各単独効果およびその交互作用(同 9 位)の和に関する SHAP Dependence Plot を示す。ビリルビン 1.2mg/dl 以上での単独作用項による癌リスク低下は、20 代では交互作用項によりほとんど打ち消されている。一方、年代が上がるにつれて、交互作用項の上昇は抑えられるため、ビリルビン高値(概ね 1.2mg/dl 以上)による大腸癌発症リスクの減少は、高齢者ほど特に効果が大いことが分かった。この原因として、若年層ではそもそも癌発症リスクが低く、癌発症リスクが低下する余地が少ない一方で、高齢者ほど癌発症リスクが高く、ビリルビンの強い抗酸化作用による活性酸素の減少を通じて、癌発症リスクが低下する余地が大いことが考えられる。

参考文献

- [1] Scott M Lundberg and Su-In Lee: A Unified Approach to Interpreting Model Predictions. Proc. of Advances in Neural Information Processing Systems 30, 4768–4777. 2017
- [2] Yasunobu Nohara, Koutarou Matsumoto, Hidehisa Soejima, Naoki Nakashima: Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Computer Methods and Programs in Biomedicine, Vol. 214, No. 106584, 2022
- [3] Dhamdhere, Kedar, Ashish Agarwal, and Mukund Sundararajan: The shapley taylor interaction index. arXiv 1902.05622 (2019).
- [4] Inoguchi, T., Nohara, Y., Nojiri, C. et al. Association of serum bilirubin levels with risk of cancer development and total death. Sci Rep 11, 13224 (2021).