研究論文

交互作用効果の計量化および可視化に関する考察

岩沢宏和[1]　　大塚忠義[2]

2023 年 9 月 5 日投稿
2024 年 1 月 17 日受理

Abstract

機械学習モデルは高い予測力を示すが，予測の際の説明力に欠けている．現在，機械学習モデルを解釈可能なものとするための手法として，Interpretable Machine Learning (IML) 分野の研究が行われているが，その手法はアクチュアリーがその携わる業務で機械学習モデルを活用するには不十分である．しかし，予測精度の高いモデルを将来にわたって活用しないことはアクチュアリー業務の発展を阻害する．本稿の目的は，機械学習モデルの説明力をアクチュアリー業務に活用できるレベルに引き上げるために必要な手法および指標を提案し，基礎的な考察を行うことである．

本稿では,交互作用効果を定義したうえで，その諸性質を確認し交互作用効果を計量化するための指標として，交互作用効果割合，交互作用効果量，未解釈割合，交互作用効果重要度を提案した．そして，既存手法である PD と ALE は主効果，交互作用効果のいずれもうまく捉えられていないことを明らかにしたうえで，これらに替わる IML 手法として CE，TCE，SI を提案し，それらの指標を用いて既存の手法および代替手法の特徴を明らかにした．加えて，既存の手法と代替手法を単純なモデルに適用して比較分析することで，既存手法の問題点と代替手法の可能性を示した．また，これらの検討結果を踏まえ，アクチュアリーがその実務に活用するために，主効果項のみを使ってできるだけ多くの部分の解釈ができる手法を活用することを提案する．

Keywords : 交互作用 (Interaction effect)，IML，関数分解 (functional decomposition)，ALE，PD

1 はじめに

機械学習モデルをはじめとするデータサイエンス技術は，驚異的な速度で発展を続けている．交互作用効果が強いデータに対して機械学習モデルは高い予測力を示しており，その予測精度は伝統的な回帰分析等の手法に基づく予測とは比較にならない．一方で，機械学習モデルはブラックボックスモデルになりがちで，その内容を解釈することは困難である．そもそも機械学習モデルの主たるユーザーのニーズは予測精度の高さにあり，予測結果に対する高い説明力は求められてこなかった．一方で，機械学習モデルを解釈可能なものとするための技術に対する要望もある．機械学習モデルが捉えた交互作用効果を可視化・計量化することによって予測結果の解釈に役立てる手法として，Interpretable Machine Learning (IML) 分野の研究が行われている．特に，交互作用効果を可視化する方法としては Partial Dependence (PD) と Accumulated Local Effects (ALE) が知られている．PD にはさまざまな欠点があ

[1] 早稲田大学大学院会計研究科　東京都新宿区西早稲田 1-6-1, Email: iwahiro@bb.mbn.or.jp
[2] 早稲田大学大学院会計研究科　東京都新宿区西早稲田 1-6-1, Email: otsukata@waseda.jp