

モデル依存・非依存型の変数重要度の理論と応用

中村 知繁* 白石 博†

2025 年 3 月 31 日投稿

概要

本論文では、機械学習モデルに対する解釈可能性の向上を目的とする変数重要度の理論と応用についてレビューを行う。変数重要度の理論は、近年急速に発展しており、モデル依存型・モデル非依存型の様々な変数重要度が提案され、理論的な結果が次々に報告されている。モデル非依存型の変数重要度は、セミパラメトリック推論の枠組みによる解析や、False discovery rate を制御しつつ予測に有効な変数集合を推定する Model-X knockoffs、またはシャープレー値に対する理論的な解析が行われている。一方、従来から実務等でよく利用されてきたランダムフォレストにおける変数重要度についても、その推定においてランダムフォレストの学習過程に起因するバイアス除去の方法や、変数間の相関がある場合の変数重要度の性質などが明らかになっている。これら以外にも、変数重要度に因果的な視点を加える方法や、機械学習の解釈可能性の向上につながる Counterfactual Explanation など新しい概念が登場している。本論文の目的は、これらの近年の変数重要度および機械学習の解釈可能性の向上に関係する手法についてレビューを行うことで、理論研究者と実務家の双方に有用な知見を提供することである。

キーワード： 変数重要度, 機械学習, 解釈可能性

1 研究背景と目的

ランダムフォレスト、勾配ブースティング決定木、ニューラルネットワークなどをはじめとする機械学習モデルを実務で利用する場合には、予測精度を示す一方で、「ブラックボックス」と見なされ結果の解釈が難しいというトレードオフが存在する ([7, 52, 44])。特に、モデルの解釈が十分に行えない場合には、意思決定プロセスが透明化できず、単に技術的な課題にとどまらず、実用面での重大な制約となりうる。例えば、モデルの判断根拠が理解できなければ、その予測結果を実際の意思決定に活用することへの信頼性が損なわれるし、モデルの欠陥や偏りを発見し修正するためには、その内部メカニズムを理解することが不可欠である ([16])。さらに、専門家の知識とモデルの予測を統合し、新たな知見を得るためにも、モデルが「なぜ」そのような予測を行ったかを説明できることは重要である ([52])。特に金融リスク評価や医療診断支援などの重要な応用領域では、機械学習の活用が増加するにつれて、モデルの判断根拠を理解し説明する必要性が強く認識されるようになっていく。金融分野では、融資審査や投資判断において機械学習モデルの利用が拡大しているが、その判断理由を説明できなければ規制上の要件を満たせない場合が多い ([6, 68])。同様に医療分野でも、診断支援や

* 順天堂大学 健康データサイエンス学部 E-mail: t.nakamura.gs@juntendo.ac.jp

† 慶應義塾大学 理工学部 E-mail: shiraishi@math.keio.ac.jp