

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 17 日現在

機関番号：17401

研究種目：基盤研究(C)（一般）

研究期間：2020～2023

課題番号：20K11938

研究課題名（和文）シャプレー値を用いた機械学習結果の統計学的解釈手法の確立

研究課題名（英文）A Study on Statistical Interpretation Methods for Machine Learning Results Using Shapley Values

研究代表者

野原 康伸（NOHARA, Yasunobu）

熊本大学・大学院先端科学研究部（工）・准教授

研究者番号：30624829

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：近年、深層学習をはじめとする機械学習技術が注目され導入が進められつつあるが、なぜそのような結果が得られたかの説明・解釈性が強く求められている。本研究では、経済学分野で用いられている「複数人が協同した場合の利益の公平分配方法であるシャプレー値」を応用して機械学習モデルの解釈手法を改善する研究である。まず、モデルをどれだけ正確に解釈できたかという説明性を定量的に評価する手法を提案した。そして、説明変数同士の相関がない場合においては、各説明変数のシャプレー値の分散が大きい順に、説明変数を選択していくことで、使用できる説明変数の数が制限された場合における説明性を最大化できることを理論的に証明した。

研究成果の学術的意義や社会的意義

既存の機械学習の解釈手法において、どの説明変数が重要であるかを表す指標である変数重要度は、経験的に使われてきたものであり、理論的な裏付けはなかった。本研究では、モデルをどれだけ正確に解釈できたかという説明性を定量的に評価する手法を提案し、その説明性を最大化するという理論的な裏付けがある手法を提案した点に大きな学術的意義を有する。

近年、機械学習は様々な分野で用いられようとしている。特に、病気の診断や自動運転等、間違いが重大な結果をもたらす分野において、機械学習がなぜそのような結果を出力したかを説明することは重要である。機械学習を広く社会へ適用するにあたって、本研究の社会的意義は大きい。

研究成果の概要（英文）：In recent years, machine learning technologies, including deep learning, have been gaining attention and are increasingly being implemented. However, there is a strong demand for the explanation and interpretability of the results these technologies produce. This study applies the Shapley value; a method of fair profit distribution among multiple collaborators used in economics; to the research of interpretability methods for machine learning models. First, we propose a method to quantitatively evaluate interpretability based on how accurately the model can be interpreted. Then, we theoretically prove that, in the absence of correlation among features, selecting features in descending order of the variance of their Shapley values maximizes interpretability when the number of usable features is limited.

研究分野：機械学習

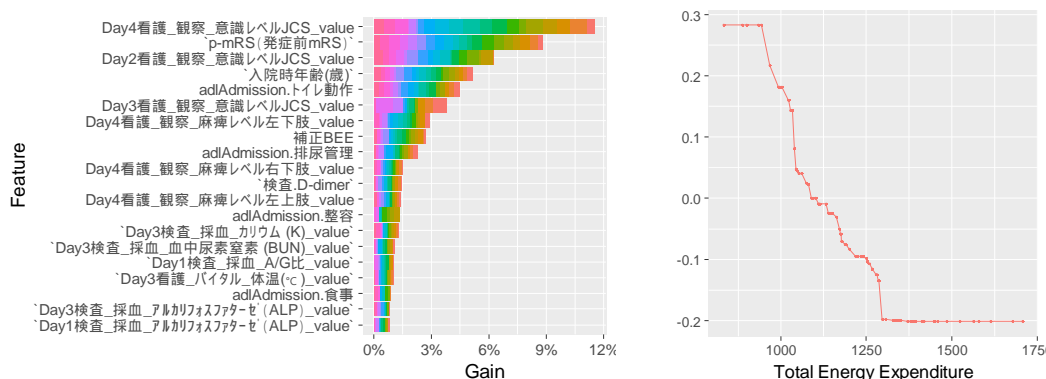
キーワード：機械学習 解釈手法 シャプレー値 説明性の定量化 変数重要度

1. 研究開始当初の背景

教師あり機械学習とは、すでにある入力(説明変数)と出力(アウトカム)のデータペアを基にして、それにうまく当てはまるような予測モデルを構築することである。例えば、レントゲン画像と診断結果のペアを多数収集して、機械学習させることで、画像診断プログラムを構築できる。ディープラーニング(深層学習)をはじめとして昨今の機械学習技術の進展は目覚ましく、十分な量のデータさえ与えれば、特徴量抽出、ハイパーパラメータの最適化、複数モデルの選択・統合化といった一連の作業が自動化され、高い予測精度を持つモデルが簡単に得られるようになってきている。しかしながら、意思決定に活用する際には、ただ判断が正しいことだけでは不十分である。なぜそのような判断がなされたのかについて合理的な説明・解釈ができることも強く求められている。

解釈性に関するアプローチとして、線形モデルのような最初から解釈可能なモデルを使うという方法があり、p 値や信頼区間といった統計学的な解釈手法が広く知られている。しかし、線形関係といった強い前提条件が必要なため、柔軟性に欠け、高精度の予測モデルを自動的に構築することは一般的に困難である。そこで、とまかく予測精度の高い予測モデルを構築したうえで、入力と出力の関係に注目してブラックボックス的にモデルを理解する方法が注目されている(Model-Agnostic 法)。現実世界から十分な量のデータが取得でき、そのデータを表すモデルを精度よく構築できるならば、そのモデルを理解・解釈することは、現実世界そのものを理解することにつながるため、新たな科学的手法となり得る。

申請者がこれまでに行った解析結果の一例を示す(脳梗塞患者のデータを用いて、退院時に全介助が必要となるかを予測したもの。図1参照)。まず、勾配ブースティング決定木(GBDT)と呼ばれる機械学習アルゴリズムを用いて、データにフィットするような予測モデルを構築する。続いて、変数重要度とよばれる値により、どの説明変数が予測にどの程度寄与しているかを計算する(同左。重要度が高い順に説明変数が並んでいる)。そして、重要変数とアウトカムの関係について Partial Dependence Plot(PDP)と呼ばれる手法で可視化を行う(図右。TEE が高いほど全介助が必要となるリスクが高くなること分かる)。



線形モデルと対比すると、変数重要度は標準化偏回帰係数、PDP は偏回帰係数 × 説明変数と対応しており、予測結果のかなりの部分を理解することができる。しかし、以下のような課題がある。

- 1) 似たような変数(例えば、毎日のJCS 値や部位ごとの麻痺レベル)が多数、重要変数として抽出されており、アウトカムとの関連が分かりにくい。線形モデルのように類似する説明変数を除外すると、アウトカムとの関係は分かりやすくなるが、予測精度の劣化を招く場合があり、かつ、予測モデルの再構築を行う必要がある
- 2) 類似する説明変数はすべて抽出し、その上位概念でグループ化することで階層的な理解を行いたい(例: 補正 BEE(TEE)と A/G 比は患者の栄養状態を表し、栄養状態が患者の予後に影響する)。これまでの手法では、説明変数のグループ化を行うと、予測モデルの再構築を行う必要があった。また、当該解析分野における専門知識を用いたマニュアル方式のグループ化だけでなく、データに基づく自動グループ化も行いたい
- 3) 変数重要度・PDP では、複数の説明変数間の交互作用項の存在を捉えられない
- 4) p 値や信頼区間を求めることができず、解析結果がどの程度信頼できるか分からない

特に機械学習手法の解釈を、現実世界を理解するための方法としてとらえた場合、4 番目の問題が特に重要である。データの解釈にはこれまで p 値などの統計学的な解釈手法が広く使われており、知識を持つ人も多いため、この課題を解決できればより多くの人に機械学習結果を受け入れてもらうことが可能となる。

線形性などの強い仮定を設けずに、現実世界への汎化能力を可能な限り高めて構築したブラックボックスモデルに対して、仮説検定や信頼区間、p 値といった統計的解釈手法をどのようにすれば適用できるかという点が本研究の核心をなす大きな問いである。

2. 研究の目的

本研究では、予測器をブラックボックス的に捉え、シャプレー値を活用して階層的な構造をもったモデルに分解することで、仮説検定や信頼区間といった統計学的な解釈手法を適用可能とし、機械学習結果を解釈する手法を開発することを目的とする。

経済学分野で用いられている「複数人が協同した場合の利益の公平分配方法であるシャプレー値」を応用することで、機械学習モデルのアウトカムを各説明変数の貢献度に応じて決まる値 (SHAP 値) の線形和の形で表現するという方法を Lundberg(2017)らが提案している。SHAP 値はデータ解析の結果として非常によい性質を持っており、ブラックボックスモデルを線形モデルのように扱うことができる。この点に注目して、線形性の前提を基に構築された各種統計的手法の適用を検討するアイデアは医療統計と機械学習の両方を研究してきた研究代表者に独自のもので、これまでに例がない。

また、交互作用の強い他の測定項目が存在する場合、同じ測定値であっても、各ユーザのシャプレー値が大きく変動(=分散大)するという性質を活用し、交互作用項の可視化と、類似変数の自動グループ化を行う。シャプレー値をベースとして、説明変数間の関連性を上位概念としてグループ化して抽出することで、より解釈しやすい階層的なモデルを作成する

3. 研究の方法

1) 変数のグルーピング化と階層モデルの構築

シャプレー値に基づいて、交互作用項の抽出と、類似変数の自動グループ化を行うアルゴリズムを開発する。これにより、当該分野の専門知識がなくても、理解しやすい階層的なモデルを構築することができ、試行錯誤の回数を減らした効率的な解析が可能となる。さらに、マニュアル方式のグループ化もできるようにし、当該解析分野における専門知識を用いたより精緻なモデルを構築可能とする。

2) 説明性の定量化手法の開発

シャプレー値によって機械学習モデルをどれだけ正確に解釈できたかという説明性を定量的に評価する手法を開発する。また、シャプレー値の定義に立ち戻って、説明変数に相関がある場合やない場合等様々なケースにおける貢献度の定量化を実施する。

3) 病院の実データを用いた検証と知識発見

開発した手法を病院の実データに適用して、解釈性などを検証する。病院における患者の転倒転落事故や糖尿病の合併症発症などに関する医療データ数万件を既に抽出しており、決定木ベースの機械学習アルゴリズムによる解析で、どの説明変数が予測に寄与しているかを概ね把握済みである。解析で得られた知見を基に、研究協力者と協同して転倒転落事故を減らすための看護業務改善や新たな治療方法の開発等に生かす。

4. 研究成果

1) グルーピング

変数グループ化という新たな解析手法を開発した。また、変数グループ化を用いることで予測モデルの再構築を伴うことなく、複数の説明変数がアウトカムに与える影響を調べることが可能となった。

2) 交互作用の分離

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。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。SHAP 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した。

3) 説明性の定量化手法の開発

シャプレー値によって機械学習モデルをどれだけ正確に解釈できたかという説明性を定量的に評価する手法を提案した。続いて、シャプレー値の定義に立ち戻って、説明変数に相関がある場合における貢献度の定量化を実施した。この結果、相関があっても、機械学習モデルの説明性は低下しないが、説明変数の貢献度を正しく分配できないことを定量的に示した。

4) 理論的な裏付けのある変数重要度

説明変数同士の相関がない場合においては、各説明変数のシャプレー値の分散(標準偏差)が大きい順に、説明変数を選択していくことで、使用できる説明変数の数が制限された場合における説明性を最大化できることを証明した。これまで各変数の重要性を表す指標としてシャプレー値の絶対値の和が広く用いられてきたが、理論的な裏付けはなかった。変数重要度としてシャプレー値の分散を用いることで、説明性を最大化できる事を理論的な裏付けを示せた。また、これまで提案されていた変数重要度では、線形モデルで用いられている変数重要度と不整合が出る場合があったが、提案手法により完全に整合性が取れるようになった。

実験を通した解析も実施し、既存手法と比べて説明性に優れていることが分かった。またその際、説明性の向上は既存手法と比べて大きくはなかったが、理論的な裏付けのなかった既存手法に、最適解と同程度の説明性が備わっていることが実証的に示せた。

5) 実データを用いた検証

脳梗塞軽症パスの重症度予測モデルの解析

A/G 比が重症度予測の重要度 8 位の説明変数として抽出されたほか、変数のグループ化や交互作用項の発見等で従来手法よりモデルの解釈が容易となった。また、複数病院間の診療データに対する予測モデルに対しても提案手法による解釈を行い、寝返り等が長期在院に関連する因子として抽出された。

癌発症モデルの解析

九州大学病院の癌コホートデータ(N=29080)を用いて、どのような要因の組み合わせががんの発症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。

虚血性心疾患の予後予測

低体重がリスク要因であることや、これまでの予後スコアでは使われていなかった D ダイマーや、C 反応性タンパク質、フィブリノーゲンなど検査結果が有用であることが示された。

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 10件 / うち国際共著 0件 / うちオープンアクセス 4件）

1. 著者名 Matsumoto Koutarou, Nohara Yasunobu, Sakaguchi Mikako, Takayama Yohei, Fukushige Shota, Soejima Hidehisa, Nakashima Naoki	4. 巻 13
2. 論文標題 Delirium Prediction Using Machine Learning Interpretation Method and Its Incorporation into a Clinical Workflow	5. 発行年 2023年
3. 雑誌名 Applied Sciences	6. 最初と最後の頁 1564 ~ 1564
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/app13031564	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Nishikitani Mariko, Ariyoshi Mie, Nohara Yasunobu, Umihara Junko	4. 巻 6
2. 論文標題 Communications Through Contemporary Tools of Information and Communication Technology: Cross-sectional Study Evaluating Health Among Separated Family Members	5. 発行年 2022年
3. 雑誌名 JMIR Formative Research	6. 最初と最後の頁 e34949 ~ e34949
掲載論文のDOI (デジタルオブジェクト識別子) 10.2196/34949	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Yasunobu Nohara, Koutarou Matsumoto, Hidehisa Soejima, Naoki Nakashima	4. 巻 214
2. 論文標題 Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital	5. 発行年 2022年
3. 雑誌名 Computer Methods and Programs in Biomedicine	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.cmpb.2021.106584	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Takanori Yamashita, Yoshifumi Wakata, Hideki Nakaguma, Yasunobu Nohara, Shinj Hato, Susumu Kawamura, Shuko Muraoka, Masatoshi Sugita, Mihoko Okada, Naoki Nakashima, Hidehisa Soejima	4. 巻 214
2. 論文標題 Machine Learning for Classification of Postoperative Patient Status Using Standardized Medical Data	5. 発行年 2022年
3. 雑誌名 Computer Methods and Programs in Biomedicine	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.cmpb.2021.106583	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Kimiyo Kikuchi, Yoko Sato, Rieko Izukura, Mariko Nishikitani, Kiyoko Kato, Seiichi Morokuma, Meherun Nessa, Yasunobu Nohara, Fumihiko Yokota, Ashir Ahmed, Rafiqul Islam Maruf, Naoki Nakashima	4. 巻 207
2. 論文標題 Portable Health Clinic for Sustainable Care of Mothers and Newborns in Rural Bangladesh	5. 発行年 2021年
3. 雑誌名 Computer Methods and Programs in Biomedicine	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.cmpb.2021.106156	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Toyoshi Inoguchi, Yasunobu Nohara, Chinatsu Nojiri and Naoki Nakashima	4. 巻 11
2. 論文標題 Association of serum bilirubin levels with risk of cancer development and total death	5. 発行年 2021年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-021-92442-2	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 山下 貴範, 若田 好史, 中熊 英貴, 野原 康伸, 岡田 美保子, 中島 直樹, 副島 秀久	4. 巻 41
2. 論文標題 患者状態把握を目的とした機械学習と共起有向グラフによる診療プロセス解析	5. 発行年 2021年
3. 雑誌名 医療情報学	6. 最初と最後の頁 29-37
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 山下 貴範, 若田 好史, 中熊 英貴, 野原 康伸, 岡田 美保子, 中島 直樹, 副島 秀久	4. 巻 41
2. 論文標題 患者状態把握を目的とした機械学習と共起有向グラフによる診療プロセス解析	5. 発行年 2021年
3. 雑誌名 医療情報学	6. 最初と最後の頁 29-37
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Soejima Hidehisa, Matsumoto Koutarou, Nakashima Naoki, Nohara Yasunobu, Yamashita Takanori, Machida Jiro, Nakaguma Hideki	4. 巻 e10252
2. 論文標題 A functional learning health system in Japan: Experience with processes and information infrastructure toward continuous health improvement	5. 発行年 2020年
3. 雑誌名 Learning Health Systems	6. 最初と最後の頁 1-12
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/lrh2.10252	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Matsumoto Koutarou, Nohara Yasunobu, Soejima Hidehisa, Yonehara Toshiro, Nakashima Naoki, Kamouchi Masahiro	4. 巻 51
2. 論文標題 Stroke Prognostic Scores and Data-Driven Prediction of Clinical Outcomes After Acute Ischemic Stroke	5. 発行年 2020年
3. 雑誌名 Stroke	6. 最初と最後の頁 1477 ~ 1483
掲載論文のDOI (デジタルオブジェクト識別子) 10.1161/STROKEAHA.119.027300	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計8件 (うち招待講演 0件 / うち国際学会 6件)

1. 発表者名 Yasunobu Nohara
2. 発表標題 Maximizing Accuracy of Shapley Additive Explanations with Limited Features --- A Quantification of Explanation Fidelity and Trade-off Between Fidelity and Number of Features --
3. 学会等名 IJCAI 2024 Workshop on Explainable Artificial Intelligence (国際学会)
4. 発表年 2024年

1. 発表者名 Koutarou Matsumoto, Yasunobu Nohara, Mikako Sakaguchi, Yohei Takayama, Hidehisa Soejima and Naoki Nakashima
2. 発表標題 Developing a Learning Health System for Delirium using XAI
3. 学会等名 12th Biennial Conference of the Asia Pacific Association for Medical Informatics (国際学会)
4. 発表年 2022年

1. 発表者名 Yasunobu Nohara, Toyoshi Inoguchi, Chinatsu Nojiri and Naoki Nakashima
2. 発表標題 Explanation of Machine Learning Models Using SHAP Considering Interaction Effects
3. 学会等名 2nd ICML Workshop on Interpretable Machine Learning in Healthcare (国際学会)
4. 発表年 2022年

1. 発表者名 野原 康伸, 井口 登興志, 野尻 千夏, 中島 直樹
2. 発表標題 交互作用を考慮したSHAPによる機械学習モデルの解釈
3. 学会等名 第26回日本医療情報学会春季学術大会
4. 発表年 2022年

1. 発表者名 Kazunori Oka, Takumi Hirahara, Yasunobu Nohara, Sozo Inoue, Koichi Arimura, Syoji Kobashi and Koji Iihara
2. 発表標題 Predictors of Intracerebral Hematoma Enlargement Using Brain CT Images in Emergency Medical Care
3. 学会等名 5th IEEE International Conference on Cybernetics (国際学会)
4. 発表年 2021年

1. 発表者名 野原 康伸, 井口 登興志, 野尻 千夏, 中島 直樹
2. 発表標題 交互作用を考慮したSHAPによる機械学習モデルの解釈手法の提案
3. 学会等名 第41回医療情報学連合大会
4. 発表年 2021年

1. 発表者名 Yasunobu Nohara, Koutarou Matsumoto, Hidehisa Soejima, Naoki Nakashima
2. 発表標題 Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital
3. 学会等名 11th Biennial Conference of the Asia Pacific Association for Medical Informatics (国際学会)
4. 発表年 2020年

1. 発表者名 Takanori Yamashita, Yoshifumi Wakata, Hideki Nakaguma, Yasunobu Nohara, Shinji Hato, Susumu Kawamura, Shuko Muraoka, Masatoshi Sugita, Mihoko Okada, Naoki Nakashima, Hidehisa Soejima
2. 発表標題 Machine Learning for Classification of Postoperative Patient Status Using Standardized Medical Data
3. 学会等名 11th Biennial Conference of the Asia Pacific Association for Medical Informatics (国際学会)
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	松本 晃太郎 (MATSUMOTO Kotaro) (60932217)	久留米大学・付置研究所・助教 (37104)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------