

機械学習回帰における Shapley 値の理論説明と事例紹介

白田 由香利

学習院大学経済学部経営学科 〒171-8588 東京都豊島区目白 1-5-1

E-mail: yukari.shirota AT gakushuin.ac.jp

あらまし 機械学習の回帰分析において Lundberg の SHAP とそのライブラリは広く活用されており、多様な応用分野に渡って回帰分析の解析の意味解釈に多大な貢献をしている。本稿では、SHAP の元の理論である Shapley の公式及び SHAP の理論を、グラフィクス教材を使ってビジュアルに説明する。SHAP の利点は各データの特性関数に基づき、各説明変数のターゲット変数への貢献を導出することである。本稿では事例をあげて、各データの特性に基づく説明変数の評価を示す。

キーワード Shapley 値, SHAP, 機械学習, 回帰, 特性関数

1. はじめに

機械学習の回帰分析において Lundberg の SHAP とそのライブラリは広く活用されており [1], 回帰分析の解析の意味解釈に多大な貢献をしている [2-5]. 本チュートリアルでは、SHAP の元の理論である Shapley の公式 [6] 及び SHAP の理論を、グラフィクス教材を使ってビジュアルに説明する。

経済学の分野で使われてきた Shapley の公式は、協調ゲーム理論で利潤分配のためのユニーク解を与えるものであり、機械学習の回帰を用いる全分野のアプリケーションにおいても有益な分析手段となる。

第 2 節では、回帰分析における説明変数が、オリジナルの協調ゲームにおけるプレイヤーに相当する関係を説明する。次に第 3 節では、Shapley 値における重要概念である特性関数を説明し、従来、計算困難であった特性関数を SHAP ではどのように近似したのかを述べる。第 4 節では Shapley の公式を説明し、本公式の具体的な適用を説明変数 1 個の場合、及び 2 個の場合で例示する。第 5 節では SHAP の優れた特長はその特性関数にあることを述べる。第 6 節と第 7 節では、日本の出生率の県別比較、及び、日本製造業における株価上昇率の企業比較という 2 つの回帰分析における SHAP を利用した評価結果を述べる。最後はまとめである。

著者は既に Shapley 値に関する以下のオンデマンド教材を WEB 上で公開しているので、合わせて参照して頂きたい。

- <https://shirotaabc.sakura.ne.jp/hosomichi/PDF/SHAP.pdf>
- <https://shirotaabc.sakura.ne.jp/hosomichi/PDF/SHAPjapan{1,2,3}.pdf>
- <https://shirotaabc.sakura.ne.jp/hosomichi/PDF/SHAPFAQ.pdf>

2. 説明変数による協調

Shapley 値の公式は、 n 人のプレイヤーによる協調ゲームにおいて、利益 M を得た場合、 n 人にどう配分すべきであるかという問題にユニークな解を与える。これを回帰分析において、 n 個の説明変数があり、データ i においてはターゲット値が y_i である場合、この y_i を n 個の説明変数への貢献としてどのように配分すべきであるか、という問題として捉える。

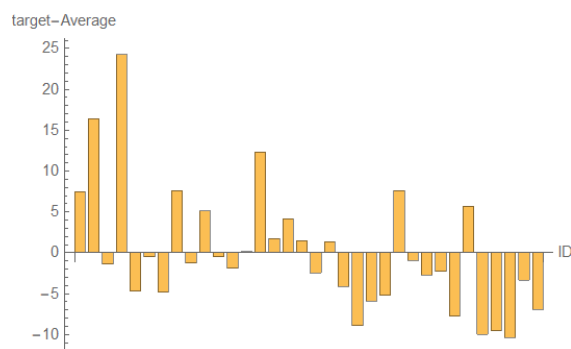


図 1: ターゲット値の平均値からのずれ(偏差)

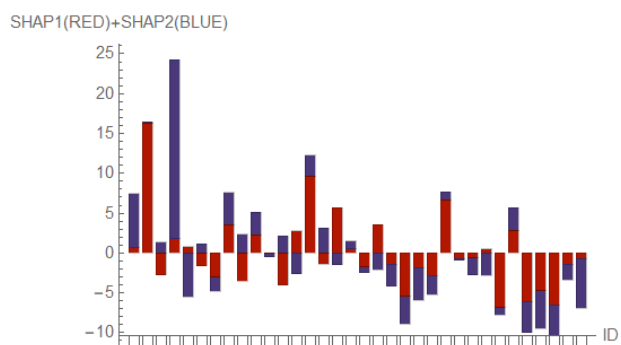


図 2: 説明変数 1 の SHAP 値(青)と説明変数 2 の SHAP 値(赤)

これをグラフィックスで説明する。簡単のため、説明変数は2個とする。図1のようなターゲット値の偏差(平均値からのずれ)があったとする。横軸はデータIDである。このターゲット値と別途与えられた説明変数値から、回帰を行い、回帰モデル $f(X)$ が得られる。その回帰モデルを用いて、各説明変数の SHAP 値が計算される。このプロセスは後述する。

各データに対して2個の SHAP 値が計算される。それを積み上げ棒グラフで表現した図が図2である。説明変数毎のターゲット値への貢献が SHAP 値である。協調ゲーム理論で複数のプレイヤーが協力して利益を得、それを分配する解を与える、ということは理解しやすいが、回帰分析でターゲット値を説明変数毎に分配する解を与える、という考え方は理解しにくいかもしれない。Lundberg の貢献は、説明変数をプレイヤーに置き換えて Shapley 値を近似的に機械学習の回帰において実現した点である。詳細は後述する。

図2に示されたように SHAP 値は負値も取る。図2に示すように、各データにおいて2個の SHAP 値の合計がほぼ図1の偏差に対応している。

SHAP の優れた点は、各データの内部構造の違いを鮮明に表してくれる点である。図2に示すように、データによって説明変数のターゲット値への貢献が大きく異なる。企業分析においては、各企業(各データ)の内部構造の違いが分かるようになる。例えば、営業利益率をターゲットとした場合、営業利益率に大きく貢献しているのは、説明変数1の売上高成長率か、あるいは説明変数2の在庫回転率か、という構造の違いが見えてくる。

経営学の企業分析においては、従来から「製造業では売上高成長率が高い企業のほうが、営業利益率が高い」という通説はあったが、売上高成長率の生の値ではその関係性は抽出困難であった(図3参照)。しかし、これを SHAP 値に変え、SHAP 値とターゲット値との関係性を見ると、高い相関性が得られる(図4参照)。図3では相関係数は0.37であったが、SHAP 値を用いると0.77に向上した。こうした製造業の売上高成長率の重要性に関する SHAP による分析は山口によって初めて行われた[7, 8]。

SHAP 値を利用することで、ターゲット値との相関係数が高くなることは理論的に言えることであり、SHAP 値の利用により従来非常に困難であった説明変数とターゲットの関係性が求められるようになった。これにより、説明変数毎の貢献値の比較が行えるようになった。

3. 特性関数

本節では、Shapley の公式を理解する上で重要な特性

関数(characteristic function)を説明する

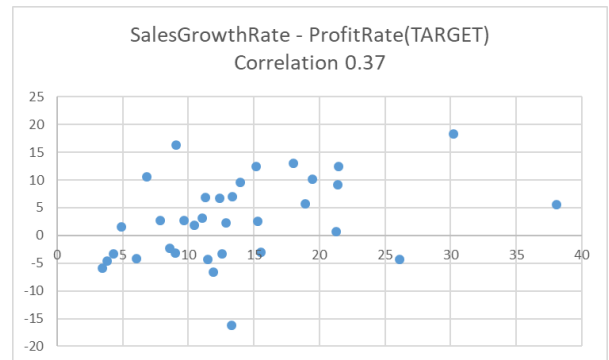


図3: 生の説明変数とターゲット値の散布図

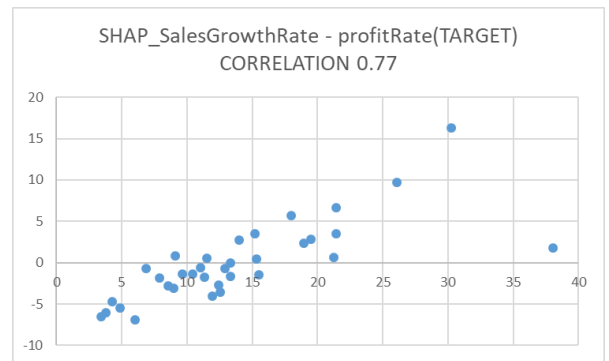


図4: 説明変数の SHAP 値とターゲット値の散布図

特性関数 v は、 n 個の説明変数の任意の部分集合に対して、その部分集合によって得られる報酬(ターゲット値)がいくらになるかを与える関数である。

$$v: 2^n \rightarrow R$$

例えば、説明変数が2個の場合、以下の4つの部分集合が存在する。

$$\{\}, \{\text{説明変数 1}\}, \{\text{説明変数 2}\}, \{\text{説明変数 1, 説明変数 2}\}$$

特性関数 v は各部分集合に対して、その利益額(報酬額)を返す。現実問題として、本特性関数を求めることは困難で、ほぼ不可能である。説明変数が2個の回帰で、説明変数1だけの貢献等計算できないからである。特性関数が定義できなければ、Shapley 値は定義できない。

この問題を Lundberg は機械学習による回帰モデル $f(X)$ を以下のように活用することで解決した[9]:
 参加していない説明変数に関しては、その説明変数の平均値を用い、 $f(X)$ を計算し、その値で v を近似する。

このような v の近似によって求められた Shapley 値を SHAP 値と呼び、以降ではオリジナルの Shapley 値と区別することにする。

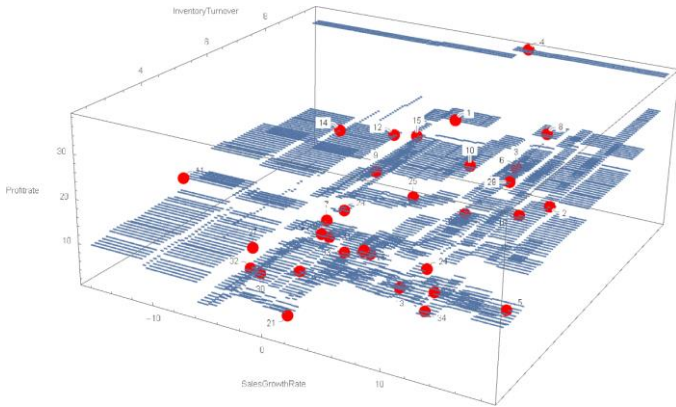


図 5: 回帰モデル $f(X)$ の例

図 5 に回帰モデル $f(X)$ の例を示した。図中、赤の点はターゲット変数の観測値を示している。この $f(X)$ を用いて、例えば、 $v(\{\text{説明変数 1}\})$ を計算したい場合、説明変数 2 には、データ全体の説明変数 2 の平均値を入れて値を求める。空集合に対する報酬額、 $v(\{\})$ を計算したい場合は、全ての引数に、平均値を入れて $f(X)$ を計算する。

4. Shapley 値の計算過程

本節では Shapley 値の公式を示し、その意味をその計算過程から説明する。

4.1 Shapley 値の公式

以下に Shapley のプレイヤー i の貢献度の期待値の解の式を与える(以下では公式と略す)。

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)]$$

以下の説明ではプレイヤーを説明変数と置き換える。 $[v(S \cup \{i\}) - v(S)]$ の項は、任意の説明変数部分集合 S に対する説明変数 i の貢献を表している。

説明変数全体集合を N とし、その要素数を $|N| = n$ で表す。次に、説明変数 i が 1 個ずつ参加して全体で協調作業をする $|N|!$ 通りの順列が等確率で起こると仮定する。

順列組合せの数から、確率を計算しよう。はじめに部分集合 S が参加したので、その順列数は $|S|!$ となる。その後説明変数 i が参加する。最後に残った部分集合が参加するが、その要素数は $(|N| - |S| - 1)$ であり、その順列数は $(|N| - |S| - 1)!$ である。よって以下が得られる。

$$|S|! \times 1 \times (|N| - |S| - 1)!$$

よって公式中の $\frac{|S|! (|N| - |S| - 1)!}{|N|!}$ の項は、初めに S が

参加して、その後 i が参加したときの確率となる。 S の取り方全てに対して合計を計算するのが、式の Σ の部分である。但し S の中には i は含まないように $N \setminus \{i\}$ とする。

次節ではこの公式を使って SHAP 値を計算するようすをグラフィクスで説明する。

4.2 説明変数が 1 個の事例

説明変数が 1 個 SGR(Sales Growth Ratio) のみの場合(単回帰)で SHAP を計算してみよう。ターゲット値は営業利益率とする。

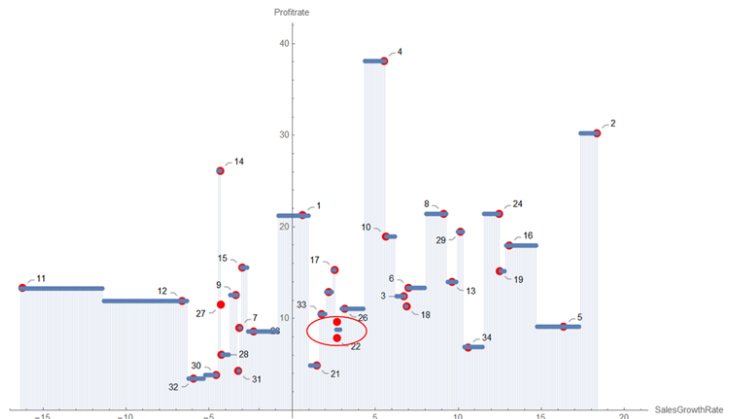


図 6: 説明変数 1 個の場合の回帰モデル $f(X)$

図 6 に機械学習によって得られた回帰モデルを示した。番号はデータ ID である。図中赤い点がターゲット値の観測値である。図 6 では、ID22 の観測値と予測値がずれている様子が見られる。

SGR の SHAP 値は以下の式で計算される。

$$\phi_{SGR} = f(SGR) - f(AVR_{SGR}) = f(SGR) - 13.8$$

部分集合 S の取り方は、 $S = \text{空集合}$ 、という 1 通りしかない。その際 $v(S) = f(AVR_{SGR})$ と近似する。

式中、説明変数 SGR の平均値 AVR_{SGR} 2.9 におけるターゲット値 $f(2.9) = 13.8$ を用いた。この値はデータが業界全体であった場合、業界平均値である。企業分析の場合、変数が業界平均値よりも上か下か、という点が非常に重要であるが、SHAP 値も正負でそれを表わしてくれる。

平均値からのずれが偏差である。この式 ϕ_{SGR} は単回帰なので、そのまま、偏差を意味する。1 変数回帰の場合、平均値からの偏差が SHAP 値となる(図 7 参照)。

回帰モデルが観測値と完全に一致しているならば、図 7 の偏差とターゲット値の散布図は 1 直線上に並ぶ。しかし図 7 を見るとずれている箇所がある。これは、ID22 のデータは初めの回帰モデルにおいて観測値と

予測値がずれていたため、SHAP 値も偏差からずれたことが確認できる(図 6 参照)。

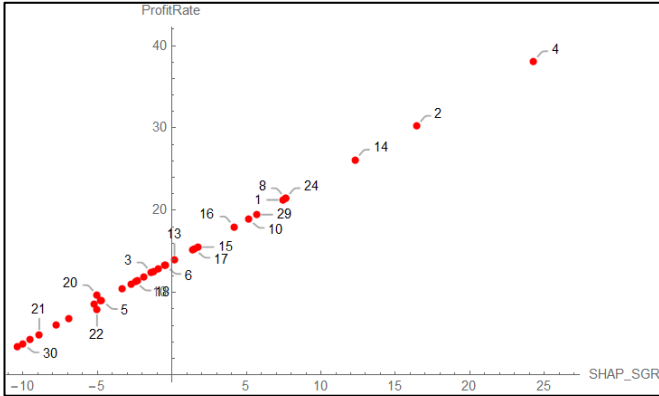


図 7: SHAP_SGR (横軸) とターゲット値 (縦軸) の散布図

4.3 説明変数が 2 個の事例

次に説明変数が 2 個 SGR と INV(Inventory Turnover Ratio, 在庫回転率)の場合を見ていく。在庫量は少ないほど良い。在庫回転率は在庫量の逆数に相当し、高いほど良い。

SGR の SHAP の式を書いてみる。以下のように、空集合を S にした場合と、{INV} を S にした場合の 2 通りの合計から構成される。その S に SGR が参加したときの貢献を計算する。2 項の確率係数は $\frac{1}{2}$ である。

$$\begin{aligned} \phi_{SGR} = & \frac{1}{2}[f(SGR, 5.5) - f(2.9, 5.5)] \\ & + \frac{1}{2}[f(SGR, INV) - f(2.9, INV)] \end{aligned}$$

本式の中で、SGR, INV の箇所に、そのデータ(企業)の説明変数の値が代入される。

INV の SHAP 式は以下のようになる。

$$\begin{aligned} \phi_{INV} = & \frac{1}{2}[f(2.9, INV) - f(2.9, 5.5)] \\ & + \frac{1}{2}[f(SGR, INV) - f(SGR, 5.5)] \end{aligned}$$

偏差を計算する際に、ベースとなる値として、 $f(2.9, 5.5)$ 及び $f(SGR, 5.5)$ が使われている。

この事例での SHAP_SGR 及び SHAP_INV を積み上げグラフで表示したものが図 2 である。

5. データの内部構造の解釈

本節では、機械学習の回帰において、回帰結果の解釈に SHAP を用いることの利点を考察する。SHAP は各データの特性関数に基づき、各説明変数のターゲット変数への貢献を導出する。特性関数による評価を行う点が利点となる。繰り返すが、特性関数はデータ毎

に異なる。例えば、企業分析では、ソニーの特性関数とパナソニックの特性関数は異なる。

特性という考え方に相反する評価方法として、絶対値による評価がある。例えば、日本企業でその収益性を示す指標である「自己資本利益率 ROE」について、2014 年に経済産業省がまとめた報告書(通称「伊藤リポート」)では日本企業の ROE で 8% を目指すことが盛り込まれた[10]。

しかし企業の内的構造は異なり、同じ高い収益性を実現するアプローチも多様である。つまり、ROE8%でも ROE のターゲットへの貢献が他の指標(有形固定資産回転率等)による貢献に比較して十分でない場合もある。反対に、ROE5%と低くとも、他の指標に比較して十分高い貢献をしている企業もあるに違いない。

全企業に一律に絶対的な値を提示し区別するよりも、各企業の特性を評価した方が、精緻な分析が行える。よって、その各企業の構造特性に適した、特性を考慮した評価が必要であり、それを実現したものが SHAP である。

医学の例で説明すると、同じ睡眠時間と食生活環境でも、人により病気の発症率が異なるということはある。理由は、個人差により、少ない睡眠時間でも健康を保つのに十分である、というような特性関数の違いが存在するからである。

現在、我々は SHAP 値を活用して、各説明変数のターゲット値への貢献を求めているが、これを可能としたのは、以下の 2 つの大きな学問的発展によるものと言える。

- (1) Shapley による Shapley 値の公式の導出
- (2) Lundberg らによる、機械学習回帰による特性関数の近似方法発案及び、Shapley 値の機械学習回帰への適応の発案

6. 事例 1: 日本の県別出生率の回帰

本節では、出生率に関する日本の近都道府県別回帰分析を行った結果を示す。著者は ICIM2020 において回帰分析の結果を発表した。しかしその時の発表においては、分析結果評価 SHAP を用いていなかった[11]。しかし、本稿のように SHAP を用いることで、各県の構造がより詳細に分析可能となった。

回帰に用いたデータを説明する。説明変数は 4 個だ。データは政府統計の総合窓口 (e-stat.go.jp) から取得した。

ターゲット変数: 出生率(100 人あたりの出生数) 2017 年データ

説明変数:

- ① 転入者数(100 人あたり) 2017 年データ

- ② 世帯サイズ(総人口÷世帯数) 2015 年データ
 - ③ 婚姻件数(100 人あたり) 2017 年データ
 - ④ 保育所等在所児数(100 人あたり) 2016 年データ (以下、 保育園と略す)
- 回帰アルゴリズムは、機械学習の XGBoost を用いた。

表 1： 東京と神奈川の比較

	転入	世帯サイズ	婚姻	保育園	出生率
東京	6.00	2.02	0.63	1.60	0.81
神奈川	4.56	2.29	0.51	1.40	0.75
	転入SHAP	世帯サイズSHAP	婚姻SHAP	保育園SHAP	出生率平均
東京	0.044650316	-0.016629068	0.06118789	0.001049976	0.09
神奈川	0.000557554	-0.000497122	0.05410064	-0.024387617	0.03

表 1 に東京都と神奈川県の説明変数の生のデータと SHAP 値の表を示した。東京都の 4 個の SHAP 値を見ると、のそ中で、世帯サイズのみが負の値となり、平均値を下回っていることが分かる。神奈川県では、保育園の SHAP 値が負値でしかも大きい。これにより、神奈川県では、保育園数の少なさが出生率へのマイナス要因として大きく働いていることが分かる。

7. 事例 2：日本企業の株価成長率

2 番目の事例として、日本の製造業 3 種の株価の上昇率の回帰分析結果を述べる。詳細は[12]を参照して頂きたい。

データは、自動車製造業(79 社)、精密機械製造業(26 社)、電気機器製造業(112 社)の 2012 年 11 月 13 日から 2015 年 8 月 18 日の期間の株価である。データは「日経 financial quest」から取得した。

ターゲット変数：株価上昇率=2015 年 8 月 18 日株価÷2012 年 11 月 13 日株価

説明変数：

- ① 売上高成長率[%]
- ② 売上高営業利益率[%]
- ③ 自己資本当期純利益率 (ROE) [%].
- ④ 使用総資本事業利益率 (ROA) [%]
- ⑤ 棚卸資産 (在庫) 回転率[回]
- ⑥ 有形固定資産回転率[回]

各説明変数の値は、10 年間(2006 年度から 2016 年度)の年次データの平均値を使っている。企業指標は年により変動があるので、1 年ではなく、少なくとも 5 年分の長期間の平均値を取る必要がある。

X 社と Y 社の SHAP 値を比較してみる。ターゲット値は X 社が 2.58 で Y 社の 1.52 よりも高い。しかし、SHAP の合計値 (ターゲット値の偏差) は両社との負の値で、業界平均を下回ることが分かる。X 社のほう

が-0.0008 なので、X 社の株価上昇率が Y 社よりも高い。個々の説明変数の SHAP 値を 2 社で比較してみると、「売上高成長率」の SHAP 値の違いが顕著である。X 社は 0.136 と高いが、Y 社は-0.806 と低い。X 社の内部構造を見ると、SHAP_売上高成長率は高く、SHAP_売上高営業利益率が-0.15 と低いことが分かる。X 社の場合、売上高は増加したが、営業利益率が低いという問題があることが読み取れる。

Y 社の SHAP 値を見ると、最大の課題は売上高成長率の低さ(-0.806)が挙げられる。次に ROE の低さ (-0.14)が課題と言える。Y 社の棚卸資産回転率及び有形固定資産回転率の SHAP は 0 に近く、他の指標に比較してターゲット値へ正方向に貢献していることが分かる。

表 2： X 社と Y 社の SHAP 値の比較

	株価上昇率	1. 売上高成長率[%]	2. 売上高営業利益率[%]	3. 自己資本当期純利益率 (ROE) [%]	4. 使用総資本事業利益率 (ROA) [%]	5. 棚卸資産 (在庫) 回転率[回]	6. 有形固定資産回転率[回]	SHAP合計
X社	2.58	0.1363	-0.1525	0.0033	0.0246	-0.0487	0.0362	-0.0008
Y社	1.52	-0.8057	-0.0794	-0.1383	-0.0472	-0.0030	0.0133	-1.0604

3 業種 (自動車製造業、精密機械製造業、電気機器製造業)の分析をした結果、いずれの業種でも、生の説明変数値に比較して、SHAP 値を用いることで、いずれの説明変数においてもターゲット値との高い相関が得られた。

SHAP を用いた企業分析の利点は、企業の内部構造を反映した分析が行える点である。それは絶対値による判定に比較して、その企業の特性を反映した評価を行っているからである。

8. まとめ

本稿では、Shapley による Shapley 値の公式及び、それを機械学習の回帰分析に適応した SHAP について解説を行った。現在、SHAP とそのライブラリは広く活用されており、多様な応用分野に渡って回帰分析の解析の意味解釈に多大な貢献をしている。その SHAP の利点は各データの特性関数に基づき、各説明変数のターゲット変数への貢献を導出することである。SHAP による評価を行う際、Shapley の公式の数式の意味を真に理解していないと、分析が浅くなってしまうと筆者は危惧する。本稿では、この数式説明のため、単回帰の場合、SHAP がターゲット値の偏差になることから説明を始めた。説明変数 2 個の場合の回帰では、2 個の SHAP 値の合計が、ターゲット値の偏差にほぼ等しくなることをグラフィックスで示した。

Lundberg の SHAP アプローチの優れた点は、回帰モデルを用いて特性関数を計算する時、参加していない説明変数の値として平均値を用いる点である。もし平

均値の箇所で回帰モデルが大きく観測値とずれていた場合、一般には、決定係数の低い回帰モデルを使った場合、SHAP 値への信頼性も減少する。回帰モデル及び SHAP 値とターゲット値の関係を本稿のように可視化で理解することができれば、SHAP による評価を行う際、どのようなチェックポイントで信頼性を確認していくか、自分独自の設計が可能になると考える。

機械学習の分析手法は益々複雑化し、その数学理論の理解は困難となっていく。しかし、理論を理解しないで分析をすることは結果の解釈の信頼性を減じさせる可能性がある。著者としては、可能な限り数式を使わずに理論を説明するグラフィクス教材を今後も作成していきたい。

謝辞

本研究は部分的に、科研 B, 20H01537(代表：白田由香利), 及び, 2021 年度学習院大学計算機センター特別プロジェクトの助成によるものである。

参考文献

- [1] S. Lundberg, and S.-I. Lee, "An unexpected unity among methods for interpreting model predictions," *arXiv preprint arXiv:1611.07478*, 2016.
- [2] C. Yang, M. Chen, and Q. Yuan, "The application of XGBoost and SHAP to examining the factors in freight truck-related crashes: An exploratory analysis," *Accident Analysis & Prevention*, vol. 158, pp. 106153, 2021.
- [3] O. Sagi, and L. Rokach, "Approximating XGBoost with an interpretable decision tree," *Information Sciences*, vol. 572, pp. 522-542, 2021.
- [4] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: an interpretation framework using XGBoost and SHAP values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466-490, 2021.
- [5] S. B. Jabeur, S. Mefteh-Wali, and J.-L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Annals of Operations Research*, pp. 1-21, 2021.
- [6] L. S. Shapley, "A value for n-person games," *Contributions to the Theory of Games*, vol. 2, no. 28, pp. 307-317, 1953.
- [7] K. Yamaguchi, "Intrinsic Meaning of Shapley Values in Regression," 2020 11th International Conference on Awareness Science and Technology (iCAST), IEEE, pp. 1-6, 2020.
- [8] K. Yamaguchi, "Feature Importance Analysis in Global Manufacturing Industry," *International Journal of Trade, Economics and Finance (IJTEF)*, vol. 13, no. 2, pp. 28-35, 2021.
- [9] S. M. Lundberg, and S.-I. Lee, "A unified approach to interpreting model predictions." pp. 4765-4774.
- [10] 日本経済新聞, "ROE とは 日本企業、8%目標," *日本経済新聞 WEB 版*, pp. <https://www.nikkei.com/article/DGXZQODL1637A0W1A210C2000000/>, 2021/02/16, 2016
- [11] Y. Shiota, and K. Yamaguchi, "Finding Dominant Factor That Affects Crude Birth Rates in Japanese Prefectures." pp. 73-77.
- [12] Y. Shiota, M. Fujimaki, E. Tsujiura, M. Morita, and J. A. D. Machuca, "A SHAP Value-Based Approach to Stock Price Evaluation of Manufacturing Companies." pp. 75-78.