解釈可能な機械学習モデルの金融データへの適用: 協力ゲーム理論を用いた新たな手法の開発と実証分析

篠 潤之介 *

概要

本分析では, 高度に複雑な機械学習モデルを用いて計算された金融・経済データの予測値を, 人間が解釈可能な形に要因分解する Additive Feature Attribution（AFA）について概説して, 実際の金融・経済データへの適用可能性を検討する（AFA とは, 例えば, 3 つの変数 X・Y・Z を用いて資産価格を機械学習モデルで予測する際, 予測値の「どの部分が X によるものなのか」「どの部分が Y によるものなのか」「どの部分が Z によるものなのか」を分解する手法である）.

具体的には, まず, AFA の代表的な手法である SHAP と, Hiraki, Ishihara and Shino [6] で提示された SHAP の代替的な手法について, その特徴や理論的な背景レビューし, 手法間の相違を議論する. 次に, それらの手法を協力ゲームの数値例および実際のデータ (金価格および有効求人倍率) に適用して, 手法間の違いがどの程度の実際の要因分解のパターンの違いをもたらすのかを分析する. 分析の結果, グラフを用いた視覚的な比較からは, (I) 協力ゲーム理論の解概念である残余均等配分解をベースにした AFA については他の AFA との違いが確認できた一方, (II) それ以外の AFA の間では, 視覚的にはっきりと確認できるほどの大きな差異はみられなかった. 一方で, (III) 協力ゲームの数値例における分析や, 異なる AFA 間の値の差の絶対値をベースにした比較からは, AFA の間で分解パターンに相応の違いが生じること, またそうした AFA 間の分解パターンの違いは, カーネルの形状をある程度反映したものとなっているということ, が明らかになった.

* 早稲田大学国際学術院. junnosuke.shino@waseda.jp

1　はじめに

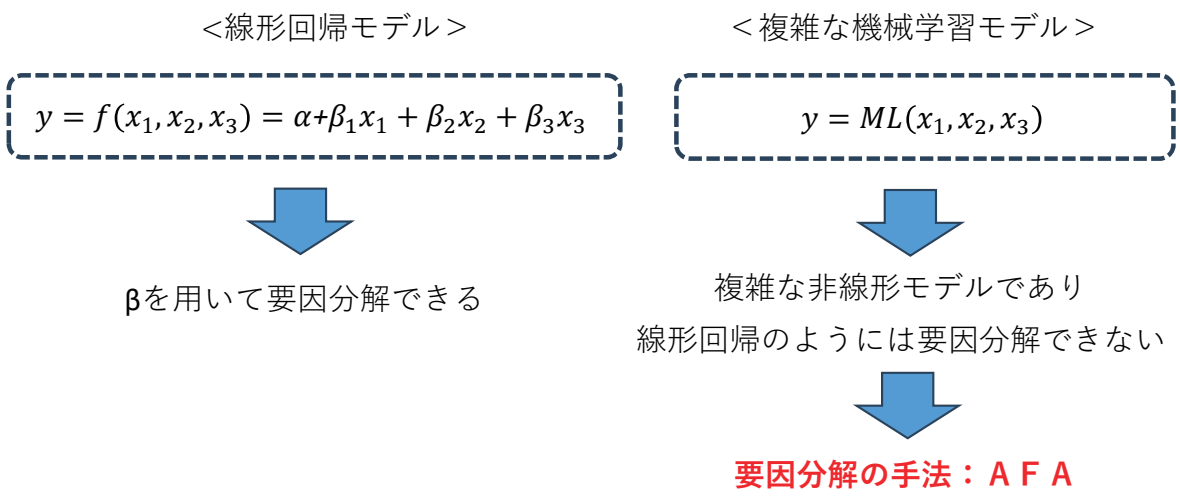
近年, 人工知能（Artificial Intelligence, AI）や機械学習（Machine Learning, ML）の進展により, 経済・金融分野におけるデータ分析の手法が大きく変化している. とりわけ, 大規模かつ高次元のデータを扱う場面において, 機械学習モデルは従来の統計的手法を凌駕する予測性能を示しており, 資産価格の予測, 信用リスクの評価, マクロ経済指標の予測といった多様な応用が進んでいる. 複雑で非線形な関係性を捉えることができる点は, 機械学習の大きな強みであり, 金融実務や政策分析における重要なツールとなりつつある.

一方で, これらの高度なモデルは内部構造が複雑であるため,「ブラックボックス」であるという批判を受けることが少なくない. 特に金融・経済分野においては, モデルが導出する結果の根拠や判断基準を明示することが強く求められる. 予測精度が高いだけでは不十分であり, モデルの出力がどのような要因に基づくのかを理解・説明可能であることが, 実務的・倫理的な観点から不可欠である. このような背景のもと, 近年注目されているのが, 機械学習モデルによる意思決定や予測を解釈しやすくするための手法を指す「説明可能な AI（Explainable AI, XAI)」のアプローチである. 本研究が着目する SHAP（SHapley Additive exPlanations）は, 協力ゲーム理論に基づく一貫性のある手法であり, 個々の特徴量が予測に与える寄与度を定量的に評価することが可能である.

より具体的には, XAI のうち, AFA（Additive Feature Attribution）とは, 複雑な機械学習モデルの予測値を個々の特徴量の貢献度 (寄与度) に分解することで, 各特徴量が予測に与える影響を定量化・可視化する手法である. AFA とは, 例えば, 3 つの変数 X・Y・Z を用いて資産価格を機械学習モデルで予測する際, 予測値の「どの部分が X によるものなのか」「どの部分が Y によるものなのか」「どの部分が Z によるものなのか」を分解する手法である. 予測モデルが線形回帰モデルであれば, 推計されたパラメータを用いて要因分解を行うことができる. しかし, ニューラルネットワークやアンサンブルツリーといった「複雑な」機械学習モデルが予測モデルである場合, そうした回帰パラメータを用いた要因分解を行うことはできない (図 1).

AFA の具体的な手法である SHAP は, 協力ゲーム理論の解概念であるシャープレイ値

図 1: 回帰分析におけるパラメータを用いた要因分解と機械学習モデルにおける AFA を用いた要因分解



(Shapley [16]) に基づく AFA であり, Lundberg and Lee [11] (以下「LL 論文」と呼ぶことにする) によって定式化されて以降, 近年, 機械学習や AI の分野において, 急速に分析・研究が進められている.[*1] 例えば, 計算コストの観点からは, SHAP の計算速度を短縮化するための TreeExplaner (Lundberg et al. [10]) や Fast SHAP (Jethani et al. [8]) といった手法が開発されている. 実際のデータを用いた SHAP の適用としては, 医療やヘルスケアの分野を中心として, 様々な分野で分析が蓄積されつつある.[*2]

さらに, ここ数年の間で, 経済・金融関連のデータに対して, SHAP を用いて機械学習モデルを解釈可能な形で適用・分析する研究が展開されつつある. Jabeur et al. [7] は, 金価格を 6 つの機械学習モデルを用いて予測した後, 各予測について SHAP を適用して比較分析を行い, XGBoost モデルとそれに対する SHAP の適用が分析上有効であることを主張した. 英国中央銀行 (BOE) のワーキングペーパーとして公表された Buckmann and Joseph [2] は, 米国の失業率を対象に, 機械学習モデルの予測精度の比較評価, SHAP による予測値の要因分解,

[*1] Lundberg and Lee [11] (LL 論文) は, 2017 年の NeurIPS（Conference on Neural Information Processing Systems, 機械学習や人工知能（AI）分野で最も権威のある国際会議のひとつ）に掲載されたプロシーディングであるが, 同論文の引用件数は 2025 年 6 月時点で 3,6000 件を超えており, XAI におけるもっとも基礎的な文献の 1 つとなっている.

[*2] 医療分野における XAI の活用状況を体系的にレビューし, 特に SHAP や LIME (後述) などの手法が診断支援や疾患予測モデルの解釈性向上に貢献していることを示した Loh et al. [9] や, 全身麻酔中の際の低酸素血症の予測に SHAP を適用した Lundberg et al.[12] などが挙げられる.

変数間の非線形関係の可視化, SHAP の統計的な検証 (Shaprey Regression) 等の, SHAP を中核とする機械学習モデルを用いた分析ワークフローを示し, 政策当局の実体経済分析や局面判断において, AFA ないし SHAP を活用することの有効性を示した.[*3] 一方, 欧州中央銀行 (ECB) のワーキングペーパーとして公表された Bluwstein et al. [1] は, SHAP を用いて金融危機の予測に有用な金融経済変数を特定し, 可視化を行った.[*4] わが国においても, 鷲見 [20] は, SHAP を用いて通貨オプション市場における投資家センチメントを分析し, その主要な変動要因が金融ストレス指数や米国イールドカーブであることを明らかにした. 森ほか [19] は, 125 か国の新型コロナウィルス新規感染者数および 36 種類の特徴量からなるパネルデータにランダムフォレストモデルを適用し, SHAP を用いて各特徴量の重要度を計測した.

本稿では, SHAP およびその代替的な手法について包括的にレビューし, それらを単純化したゲームや実際の金融・経済データに適用し, 比較分析を行う. 具体的には, まず, 既存の AFA の代表的な手法である SHAP と, Hiraki, Ishihara and Shino [6] (以下「HIS 論文」と呼ぶことにする) によって提示された SHAP と代替的な複数の手法について概説する. 特に, LL 論文をベースに, AFA の基本的な別の手法である, LIME（Local Interpretable Model-agnostic Explanations, [13]）およびそのカーネルとの関係性に着目して, これらの手法の比較を行う. 次に, Jabeur et al. [7] 等に基づき, これらの複数の手法を商品価格 (金価格)および有効求人倍率に適用して, 異なる手法の間でどの程度の違いが生じるのか, 比較分析を行う. そして, Hiraki, Ishihara and Shino [6] が提示した手法を活用するためのポイントや,今後の分析の方向性について議論する.

本論文の次節以降の構成は以下の通りである. 2 節では SHAP およびその代替的手法についてレビューする. 3 節では, 協力ゲームの数値例を用いた SHAP とその代替的手法の比較分析を行う. 4 節では, これらの手法を金価格の時系列データに適用して比較を行う. 5 節はまとめと結論である.

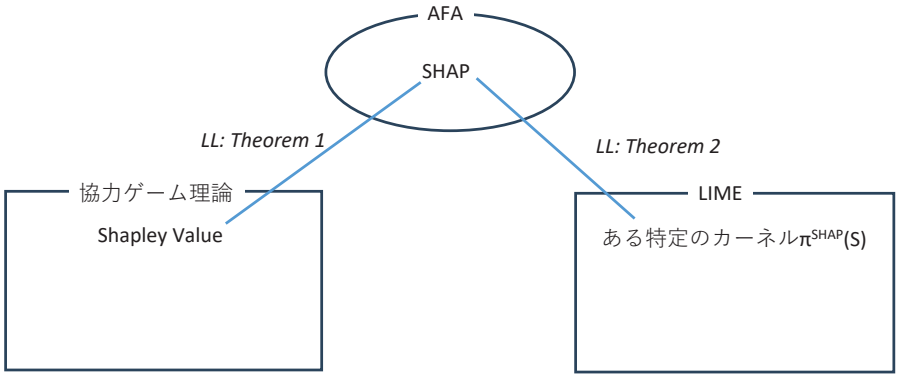
[*3] 後に, 中央銀行とその政策を専門的に扱う *International Journal of Central Banking* 誌に掲載された.
[*4] 後に, *Journal of International Economics* 誌に掲載された.

2　SHAP とその代替的手法についてのレビュー

2.1　LL 論文における SHAP と HIS 論文におけるその代替的手法に関する議論の全体感

AFA の具体的手法として SHAP を提示した LL 論文では, SHAP を 2 つの観点か特徴づけている (図 2). 1 つめの観点は, 協力ゲームの解概念であるシャープレイ値を, AFA の文脈に適用したものとして SHAP を特徴づけるものである (LL 論文の Theorem 1). もう 1 つの観点は, AFA の別の基本的な手法の 1 つである, LIME の具体的な定式化として SHAP を特徴づけるものである. 特に, LIME を定式化する際に必要となるカーネル関数 (図 2 における $\pi^{SHAP}(S)$. 詳細は 2.4 節で解説) を特定の形に限定することで, それが SHAP と一致することを示している (同論文の Theorem2).

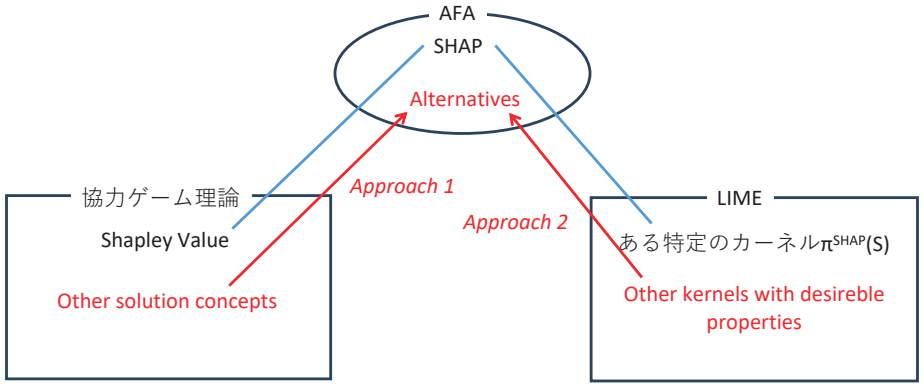
図 2: LL 論文における SHAP の特徴づけ



一方, HIS 論文では, この LL 論文の分析のフレームワークを踏襲しつつ, それぞれの観点から, SHAP の代替案を提示している (図 3). まず, 1 つめの「協力ゲームの解概念としての SHAP」について, HIS 論文では, 協力ゲーム理論における解概念はシャープレイ値の他にも多くのものがあり, それらを同様に AFA として定式化し, それを SHAP と比較することの重要性を指摘した. そのうえで, 協力ゲーム理論における残余均等配分解と最小二乗プレ仁を用いて, SHAP と代替的な AFA を導出・提示した (図 3 における左側の矢印). 次に, 2 つめ

の「LIME において特定のカーネル関数を仮定することで導出される SHAP」について, HIS 論文では, この SHAP に特定のカーネル関数が, カーネル関数が満たすべき性質を満たしていないことを指摘した. そのうえで, この性質を満たすカーネル関数を定義して, そこから SHAP と代替的な AFA を導出・提示した (図 3 における右側の矢印).

図 3: **HIS** 論文における代替的手法の提示



以上が LL 論文と HIS 論文の全体感である. 2.3 節でより詳細なレビューを行う前に, 次の 2.2 節では, LL 論文における SHAP と, HIS 論文で提示された様々な代替的手法のうちもっともシンプルな, 協力ゲーム理論の解概念である残余均等配分解を用いた AFA (残余均等配分 (Equal Surplus) の頭文字をとって, SHAP に対して ES と呼ぶことにする) について, 具体例を用いながらそのイメージを把握することにする.

2.2 簡単な比較: SHAP と ES (残余均等配分を用いた AFA) の違い

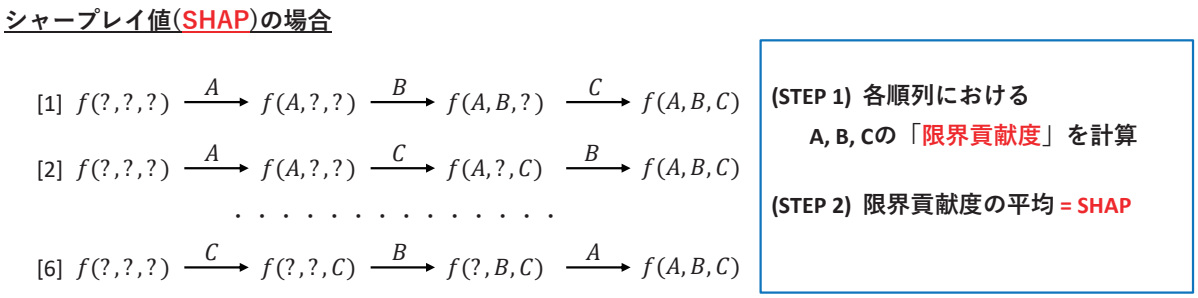
ここでは具体例を用いて SHAP と ES の違いを把握する. 学習済の機械学習モデルを f, 特徴量は A, B, C の 3 つであるとする. 例えば, 株価リターンを予測するとして, 機械学習モデル f はランダムフォレストやニューラルネットといった「複雑な」モデル, 特徴量は計量分析における「独立変数」「説明変数」のことであり, 企業収益, 配当性向, 為替レートといった変数が特徴量の候補となる.

AFA が解くべき問題は, ある観測値において, 特徴量 A, B, C の値が全て既知である場合の予測値 $f(A, B, C)$ と, 特徴量 A, B, C がすべて未知である場合の予測値 – これを $f(\emptyset)$ と

する − の差, すなわち $f(A,B,C)-f(\emptyset)$, を, 特徴量 A,B,C に配分する手法のことである. $f(A,B,C)-f(\emptyset)$ は, すべての特徴量が既知となったときの「予測の改善度」とみなすことができる. これを A,B,C の「予測の貢献度」に応じて配分する手法が AFA である. 協力ゲーム理論においては $f(A,B,C)-f(\emptyset)$ はいわゆるプレイヤー間で配分される「パイの大きさ」にあたり, その配分方法として, シャープレイ値をはじめとする様々な解概念が提示されてきた.

それでは, まず SHAP について見てみよう (図 4). SHAP では, まず, 全ての特徴量が未知の状態である $f(\emptyset)$ (図 4 では $f(?,?,?)$ と表示) からスタートし, 1 つずつ特徴量が加わる状況を想定する. 図中で [1] とあるケースでは, まず特徴量 A が既知になる. ここで A の予測の「限界貢献度 (marginal contribution)」を $f(A)-f(\emptyset)$ とする ($f(A)$ は図中では $f(A,?,?)$ と表示). 次に, 特徴量 B が既知となる. ここで B の限界貢献度は, B が既知となる前の予測値 $f(A)$ と, B が既知となったときの予測値 $f(A,B)$ の変化幅 $f(A,B)-f(A)$ とする. 最後に, 特徴量 C が既知となったときの C の限界貢献度を $f(A,B,C)-f(A,B)$ とする.

図 4: **SHAP**（シャープレイ値を用いた **AFA**）の計算方法

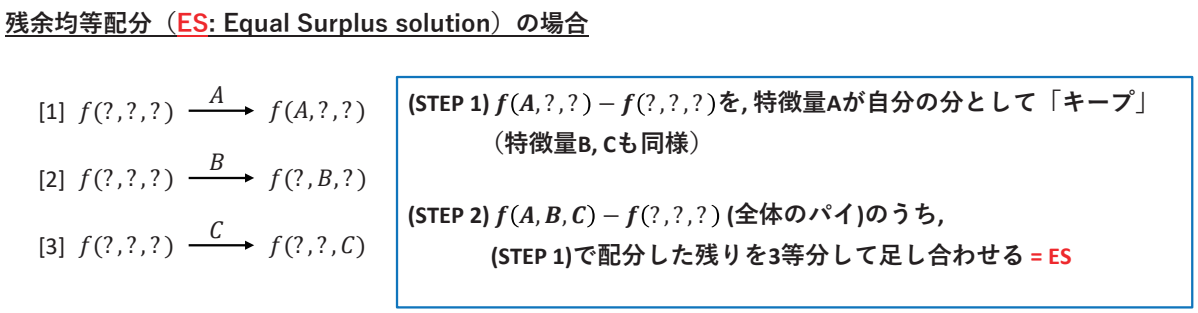


以上でケース [1] における A, B, C の限界貢献度が算出できた. ケース [1] はいわば順列 $A\to B\to C$ に対応するため, 1 つずつ特徴量が加わっていくケースの総数は A, B, C の順列の場合の数, $3!=6$ 通りある (図 4 では [1] ,[2], ..., [6] と表示されている). それぞれのケースにおいて A, B, C それぞれの限界貢献値を計算し, それをケースの総数 6 で割った, いわば「限界貢献度の平均」が SHAP となる. LL 論文ではこのシャープレイ値の考えに則って SHAP が定義された.

次に, 残余均等配分解 (ES) についてみてみよう (図 5). ES でも, まず, 全ての特徴量が

未知の状態である $f(\emptyset)$ からスタートする．しかし, 3 つの特徴量全てが加わる順列を考えるのではなく, 考慮するのは $f(\emptyset)$ から 1 つめの特徴量が加わる状態のみである．．図中で [1] とあるケースにおいては, すべての特徴量が未知である場合の予測値 $f(\emptyset)$ から, 特徴量 A が既知になったときの予測値 $f(A)$ の変化幅 $f(A) - f(\emptyset)$ を, SHAP 同様に A の予測の「限界貢献度 (marginal contribution)」とする．ES では, これを特徴量 A が自分の取り分として「キープ」すると考える．同様に, B は $f(B) - f(\emptyset)$, C は $f(C) - f(\emptyset)$ を自分の取り分として「キープ」する．この各特徴量が自分の取り分をキープする段階がいわば Step 1 である．Step 2 においては,「全体のパイの大きさ」である $f(A, B, C) - f(\emptyset)$ から, Step 1 において各特徴量がキープした分を差し引いた「残余」を, 3 つ特徴量で均等配分し, Step 1 のキープした分に加える．これが ES=残余均等配分に基づく AFA である．

図 5: **ES** (残余均等配分解を用いた **AFA**) の計算方法



SHAP と ES を比べると, 当然, SHAP の方がより多くの情報を用いて計算されている．すなわち, 上記の例において, SHAP は特徴量が 2 つが既知の場合における予測値である $f(A, B)$, $f(A, C)$ および $f(B, C)$ の情報を考慮して算出されるが, ES はこの情報を考慮しない．したがって,「各特徴量の予測の貢献度に応じて配分する」AFA として, SHAP は ES に比べるとより「フェア」な方法であると言える．また, 特徴量の数が大きくなるほど,「SHAP では考慮しているが ES では考慮していない情報量」は大きくなることから, 両者の違いは大きくなっていくと考えられる．

一方で, ES にはメリットもある．情報量についての議論といわばコインの裏表の関係であるが, SHAP の計算においては, n 個の特徴量があるとき, 計算すべき予測値の数は 2^n 個であることから, 特徴量の数が増えると SHAP の計算コストは指数関数的に増大していく．一

方で, ES においては, 計算すべき予測値の数は $n+2$ 個であることから, SHAP の計算コストの増加ペースは線形なものにとどまる. したがって, 両者の計算コストは特徴量の数が大きくなるほど拡大していく. 機械学習においては, 数多くの特徴量を扱うケースがむしろ一般的であることから, これは ES の大きなメリットであると言える. 1 節でも言及したが, SHAP は AFA として様々なメリットを持つ一方で, 計算コストの大きさが最大の問題とみなされてきた. このため, 相対的に少ない計算コストで近似的に SHAP を計算する, TreeSHAP や FastSHAP などの手法が提案されてきたが, これらの手法を用いてもなお計算コストの引き下げは限定的なものにとどまっている. HIS 論文では, この点も踏まえて, 計算コストを大幅に抑制することのできる ES を, SHAP の代替的な手法のうちの 1 つとして提示した.

SHAP と ES についての直感的な議論は以上である. HIS 論文において提示された他の代替的な手法をレビューするためには, いくつかのノーテーションを導入する必要がある. 2.3 節以降では, その準備を行ったあと, LL 論文と HIS 論文における議論をもう少し掘り下げてレビューしていく.

2.3 より正確なレビューのための準備

観測値を t 個, 特徴量の数を n 個とし ($N=\{1,...,n\}$ および $T=\{1,...,t\}$), 特徴量のベクトルを $t\times n$ 次元ベクトル $X=(X_1,...X_j,...,X_n)$ とする. f を学習済モデル, $Y=(y_1,...,y_t)'$ を f による予測値とする ($Y=f(X)$).

N のべき集合の要素 $S\in 2^N$(協力ゲーム理論では提携と呼ぶ) に対し, $x_{\tau,S}=\{x_{\tau,j}|j\in S\}$ とする. $x_{\tau,S}$ は, τ 番目の観測値における S に含まれる特徴量からなるベクトルである. また, $X_S=\{X_j|j\in S\}$ とする.

協力ゲーム理論において, 特性関数形ゲームは (N,v) で表現される. $N=\{1,...,n\}$ はプレイヤーの集合, v はべき集合 2^N 上の実数値関数である. 今, τ 番目の観測値において, 提携 S に対して実数値関数 $v_\tau:2^N\longrightarrow R$ を以下の (1) で定義すると, τ についての特性関数形ゲームが 1 つ定まる:

$$v_\tau(S)=E\left[f(x_{\tau,S},X_{N\setminus S})\right]. \tag{1}$$

$v_\tau(S)$ は,「x_τ において, S に含まれる特徴量 $x_{\tau,j}(j\in S)$ が分かっているが, S に含

まれない特徴量 $x_{\tau,k}(k \in N \setminus S)$ は未知であるときの f の予測値」である．$v_\tau(N) = E\left[f(x_{\tau,1},...,x_{\tau,n})\right] = f(x_{\tau,1},...,x_{\tau,n})$ かつ $v_\tau(\emptyset) = E[f(X_1,...,X_n)] = E[f(X)]$ である．協力ゲーム理論の分析においては，$v(\emptyset) = 0$ を仮定することが多いが，ここでは一般的に $v_\tau(\emptyset) \neq 0$ であることに留意する．

例 **1** 観測値 4 個 $(t = 4)$，特徴量 3 個 $(n = 3)$ の場合を考える．

$$X = \begin{pmatrix} x_{11}, x_{12}, x_{13} \\ x_{21}, x_{22}, x_{23} \\ x_{31}, x_{32}, x_{33} \\ x_{41}, x_{42}, x_{43} \end{pmatrix}. \tag{2}$$

4 番目の観測値 $\tau = 4$ に対応する特性関数形ゲーム v_4 は，以下の通り定まる：

$$\begin{aligned}
v_4(\emptyset) &= \frac{1}{4}\sum_{i=1}^{4} f(x_i) \qquad \text{ただし} x_i = (x_{i1}, x_{i2}, x_{i3}) \\
v_4(1) &= \frac{1}{4}\{f(x_{41}, x_{12}, x_{13}) + f(x_{41}, x_{22}, x_{23}) + f(x_{41}, x_{32}, x_{33}) + f(x_{41}, x_{42}, x_{43})\} \\
\cdots &\quad \cdots\cdots\cdots v_4(2), v_4(3)\text{も同様}\cdots\cdots\cdots \\
v_4(12) &= \frac{1}{4}\{f(x_{41}, x_{42}, x_{13}) + f(x_{41}, x_{42}, x_{23}) + f(x_{41}, x_{42}, x_{33}) + f(x_{41}, x_{42}, x_{43})\} \\
\cdots &\quad \cdots\cdots\cdots v_4(13), v_4(23)\text{も同様}\cdots\cdots\cdots \\
v_4(123) &= f(x_{41}, x_{42}, x_{43})
\end{aligned}$$

$v_4(\emptyset)$ は，「4 番目の観測値において，すべての特徴量が未知である場合の理論値」なので，x_1，x_2，x_3，x_4 が等確率で発生すると仮定し，理論値の期待値を計算する．$v_4(1)$ は，「4 番目の観測値において，1 番目の特徴量のみが既知である場合の理論値」なので，x_{41} は固定し，(x_{12}, x_{13})，(x_{22}, x_{23})，(x_{32}, x_{33})，(x_{42}, x_{43})，が等確率で発生すると仮定し，理論値の期待値を計算する．$v_4(12)$ も同様である．最後に，$v_4(123)$ は，「4 番目の観測値においてすべて特徴量が既知である場合の理論値」なので，学習済モデル f に $x_4 = (x_{41}, x_{42}, x_{43})$ を代入する．

機械学習における AFA とは，τ 番目の観測値に着目し，「すべての特徴量が既知である場合の予測値と，すべての特徴量が未知である場合の予測値の差」である $v_\tau(N) - v_\tau(\emptyset)$ を，各特徴量の貢献度 (寄与度) に応じて配分する手法である．具体的には，τ 番目の観測値に関する特性

関数形ゲーム (N, v_τ) とそこでのプレイヤー (特徴量)j に対し, 実数値関数 $\Psi_\tau(j) : N \longrightarrow R$ を考える (以後, $\Psi_\tau(j)$ を $\Psi_{\tau,j}$ と表記する). また, $\Psi_\tau = (\Psi_{\tau,1}, ..., \Psi_{\tau,n})$ とする. Ψ_τ が (3) 式を満たすとき, Ψ_τ を AFA と呼ぶ.

$$\sum_{j \in N} \Psi_{\tau,j} = v_\tau(N) - v_\tau(\emptyset). \tag{3}$$

Ψ_τ が AFA であるとき, Ψ_τ^{AFA} とも表記する.

以上の準備のもと, 2.2 節で取り上げた SHAP と ES は, それぞれ以下の (4) 式および (5) 式によって定義できる.

$$\Psi_{\tau,j}^{SHAP} = \sum_{S \subseteq N \setminus j} \frac{|S|!(n - |S| - 1)!}{n!} \left(v_\tau(S \cup \{j\}) - v_\tau(S)\right) \tag{4}$$

$$\Psi_{\tau,j}^{ES} = v_\tau(\{j\}) + \frac{(v_\tau(N) - v_\tau(\emptyset)) - \sum_{i \in N} v_\tau(\{i\})}{n} \tag{5}$$

次に, 図 3 における右側の矢印で示した, SHAP の代替的手法を定式化する際のもう 1 つのアプローチである, カーネル関数を起点とした議論を概観する.

2.4 LIME とカーネル

HIS 論文では, LIME (Local Interpretable Model-agnostic Explanations, Ribeiro et al. [13], 以下 LIME 論文と呼ぶ) におけるカーネル関数の観点から, SHAP におけるカーネル関数は,「分析対象の観測値に近い摂動サンプルほど大きなウェイトが付与されるべき」という, 本来 LIME が満たすべき条件を満たしていない点を指摘した. そのうえで, 任意のカーネル関数を用いた AFA の一般的表現を導出し, 上記条件を満たす複数の AFA を, SHAP の代替的な手法として提示した. 以下ではこの点をレビューする.

まず, LL 論文および LIME 論文の表記に従い, x が分析対象となる観測値, z は x から生成された摂動サンプル (実際上は, すべての特徴量が既知である x をベースに, 一部の特徴量を未知としたときのデータ) とする. LL 論文および LIME 論文では, 二値ベクトル z' および $z = h_x(z')$ を満たす写像 h_x を用いて z を z' に置き換えた上で分析しているが, ここでは単純化のために $x = x'$ および $z = z'$ とする. LIME 論文では,「複雑な機械学習モデル f を,

分析対象である x の近傍において,「説明可能な複雑ではない」モデル g で局所的に近似する手法」として, 以下の最小化問題を提示した:

$$\xi(x) = \underset{g \in G}{\arg\min}\, L(f, g, \pi_x) + \Omega(g).$$

ここで, $g(z) = \phi_0 + \sum_{i=1}^{n} \phi_i z_i$ (ただし $\phi_i \in R$), すなわち, 説明可能なモデルは線形であるとする. G をすべての g の集合とし, $\phi = (\phi_1, ..., \phi_n) \in R^n$ とする. π_x はカーネル関数であり, $\pi_x(z)$ で, 分析対象である観測値 x と摂動サンプル z の近接度が測られる. L は f および π_x のもとで, g が f をどの程度近似しているかどうかを測る損失関数であり, 近似度が高いほど損失が小さくなる. $\Omega(g)$ は説明可能なモデル g の複雑さを測るペナルティ項である.

以上の定式化のもと, LIME 論文では, カーネル π_x が満たすべき条件として, 以下の点を挙げている:

- x と z との近接度が高い (距離が小さい) ほど, z に付与されるカーネル (重み) は大きくなるべきである.

説明可能なモデル g は, 分析対象となる観測値 x の近くにおいて, f をより正確に近似するべきであるため, そのような g を求めるうえで, x に近い摂動サンプルを重視するように損失関数を定式化することを要請する上記の条件は, カーネル関数が当然満たすべき条件であるといえる.[*5] そして, LIME 論文では, 具体的にカーネル関数を $\pi_x(z) = exp(-D(x, z)^2)/\sigma^2$ (ただし D は距離関数, σ は散らばりの程度) と定式化している. さらに, 損失関数については, $L(f, g, \pi_x) = \sum_{z \in Z} [f(z) - g(z)]^2 \pi_x(z)$ という局所的加重二乗損失関数 (locally weighted square loss function) を想定している. すなわち, z において, 説明可能なモデルが与える予測値 $g(z)$ がそもそもの学習モデルの予測値 $f(z)$ から乖離するほど損失は大きくなる. そして, そのような損失は, z が x に近いほど (すなわち $\pi_x(z)$ が大きいほど), より重視される.

LIME 論文では, 以上の最小化問題の解を解析的に求めることはしていない. 一方, LIME

[*5] LIME 論文 [13] の 3 節を参照. 例えば, Figure 3 において, x に近い摂動サンプルは相対的に大きく表示されているが, これは当該サンプルにより大きな重みを付与していることを表している.

と SHAP の関係を考察した LL 論文では, この最小化問題に, 追加的に $\Omega(g) = 0$ の仮定を置いている. これにより, LIME の最小化問題は, 本稿の 2.3 節で導入した表記に従うと, 以下で表される[*6]:

$$\underset{\phi \in R^n}{\arg\min} \sum_{S \in 2^N} \left[\sum_{i \in S} \phi_i - \{v_\tau(S) - \phi_0)\} \right]^2 \pi_{x_\tau}(S). \tag{6}$$

さらに, LL 論文では, 説明可能なモデル $g(z) = \phi_0 + \sum_{i=1}^{n} \phi_i z_i$ に関して, いくつかの制約を課している. 1 点目は, $z = (0, ..., 0)$ のとき, $\phi_0 = f(0, ..., 0)$, すなわち, ϕ_0 は, 全ての特徴量が未知のときの学習モデルの予測値と一致していなければならない. 2 点目は, $f(x) = g(x)$, すなわち, 分析対象の観測値 x においては, $g(x)$ は $f(x)$ と一致しなければならない. この条件を局所的正確性条件 (local accuracy) または効率性条件 (efficiency) と呼び, この条件を課すことで最小化問題の解は常に AFA となる. 以上を整理すると, LL 論文の制約条件付最小化問題は, Ψ_τ^{AFA} を解とすると, 以下で表される.

$$\Psi_\tau^{AFA} = \underset{\substack{\phi \in R^n with \sum_{i \in N} \\ \phi_i = v_\tau(N) - v_\tau(\emptyset)}}{\arg\min} \sum_{S \in 2^N} \left[\sum_{i \in S} \phi_i - \{v_\tau(S) - v_\tau(\emptyset)\} \right]^2 \pi_{x_\tau}(S). \tag{7}$$

1 節でも簡単に言及したが, LL 論文は (7) 式の解が SHAP と一致するようなカーネル $\pi_{x_\tau}(S)$ が存在することを示した (Theorem 2). 一方で, HIS 論文では, 追加的な仮定 (対称性) を課すことで, (7) の最小化問題に対する解の一般的な表現を以下の通り導出した:

$$\Psi_{\tau,j}^{AFA} = \quad \phi_j \quad = \sum_{S: j \in S} \pi_{x_\tau}(S) \cdot v_\tau(S) + \frac{v_\tau(N) - v_\tau(\emptyset) - \sum_{i \in N} \left\{ \sum_{S: i \in S} \pi_{x_\tau}(S) \cdot v_\tau(S) \right\}}{n}. \tag{8}$$

(8) 式は, AFA (Ψ_τ^{AFA}) を, カーネル $\pi_{x_\tau}(S)$ の関数として表現している. このことにより, 任意のカーネルに対して AFA を求めることができ, 特定のカーネルに基づいた AFA を定式化するうえで有用である. 再び図 3 に戻ると, 右下の四角形内でカーネルを 1 つ定めると, それに応じて上部の楕円形内で SHAP の代替案=Alternative が 1 つ求まるという, 右側の赤い矢印で示された Approach 2 を, (8) 式を用いて実行することができる.

[*6] 詳細は HIS 論文参照.

以上の考察に基づき, 次節では, HIS 論文に沿って, SHAP およびその代替的な手法を定式化していく. その際に特にポイントとなるのは, 先述した「x と z との近接度が高い (距離が小さい) ほど, z に付与される重みは大きくなるべき」という条件が満たされるかどうかという点である.

2.5 カーネルに基づく AFA

まず, SHAP に対応するカーネル関数は, 以下の (9) 式である. すなわち, (9) 式を (8) 式に代入して得られる AFA は, SHAP となる:

$$\pi_{x_\tau}^{SHAP}(S) = \frac{n}{{}_nC_{|S|} \cdot |S| \cdot (n - |S|)}. \tag{9}$$

(9) 式において, $|S| = 0$ または $|S| = n$ のとき $\pi_{x_\tau}^{SHAP}(S) = \infty$ であり, 既知の特徴量の数 $|S|$ について凹型となっている. これは, LIME 論文においてカーネルが持つべき性質とされた,「z が x に近いほど, より大きな重みが付与される」すなわち π_{x_τ} が $|S|$ に関する増加関数であるべき, という条件を満たしていない.

なお, (5) で示した ES 型の AFA は, 以下のカーネル関数に基づいているが, これも $|S|$ に関する増加関数という条件を満たしてない:

$$\pi_{x_\tau}^{ES}(S) = \begin{cases} 1 & \text{if} \quad |S| = 1 \\ 0 & \text{if} \quad 2 \leq |S| \leq n. \end{cases} \tag{10}$$

すなわち, 2.2 節でも言及したように, ES は, 計算コストが相対的に小さいという大きなメリットがある一方で, カーネルが $|S|$ に関する増加関数とはなっていないという点においては, SHAP と同じようなデメリットを有しているといえる.

これに対し, HIS 論文では, さらに以下の 4 つの AFA を提示した.

2.5.1 協力ゲーム理論の解概念である LS プレ仁に基づく AFA

以下のカーネルを考える:

$$\pi_{x_\tau}^{PNucl}(S) = \frac{1}{2^{n-2}}. \tag{11}$$

これは定数, すなわち $|S|$ に関して独立なカーネル関数である. (11) 式を (8) 式に代入することで, 以下の AFA を得る:

$$\Psi_{\tau,j}^{PNucl} \quad = \phi_j = 2\left(\frac{1}{2^{n-1}}\sum_{S:j\in S} v_\tau(S)\right) + \frac{v_\tau(N) - v_\tau(\emptyset) - \sum_{i\in N}\left\{2\left(\frac{1}{2^{n-1}}\sum_{S:i\in S} v_\tau(S)\right)\right\}}{n} \tag{12}$$

$\Psi_{\tau,j}^{PNucl}$ は, 協力ゲームの分野における解概念である最小二乗プレ仁 (Ruiz et al. [14][15]) と一致することが証明できる.[*7] カーネルは定数なので,「特徴量の数 $|S|$ に関して増加関数であるべき」であるという, LIME 論文で示された条件を非常に弱い意味で満たしているといえる. そして, 再び図 3 を参照すると, 右側の Approach 2 に基づいて導出された AFA が, 実は左側の四角形である「協力ゲーム理論の世界」ではすでに最小プレ仁という解概念で提示されていたという点で, 興味深い結果であると言える.

2.5.2 線形に増加するカーネルに基づく AFA

次に, 以下のカーネル $\pi_{x_\tau}^{LnK}$ は $|S|$ に関して線形に増加しており, LIME 論文で示された条件を満たしている:

$$\pi_{x_\tau}^{LnK}(S) = \frac{|S|}{n\cdot 2^{n-3}}. \tag{13}$$

(13) を (8) に代入して, 以下の AFA を得る:

$$\Psi_{\tau,j}^{LnK} = \phi_j = \sum_{S:j\in S}\frac{|S|}{n\cdot 2^{n-3}}\cdot v_\tau(S) + \frac{v_\tau(N) - v_\tau(\emptyset) - \sum_{i\in N}\left\{\sum_{S:i\in S}\frac{|S|}{n\cdot 2^{n-3}}\cdot v_\tau(S)\right\}}{n}, \tag{14}$$

$\Psi_{\tau,j}^{LnK}$ は, カーネル上の望ましい性質を有する, SHAP と代替的な 1 つめの AFA である.

2.5.3 指数関数的に増加するカーネルに基づく AFA

以下のカーネル $\pi_{x_\tau}^{ExK}$ は $|S|$ に関して指数関数的に増加しており, それに基づく AFA である $\Psi_{\tau,j}^{ExK}$ は (16) 式の通りとなる.

$$\pi_{x_\tau}^{ExK}(S) = \frac{2^{|S|-1}}{3^{n-2}.} \tag{15}$$

[*7] 証明については著者に問い合わせされたい. また, 今後別稿で示す予定である.

$$\Psi_{\tau,j}^{ExK} = \phi_j = \sum_{S:j\in S} \frac{2^{|S|-1}}{3^{n-2}} \cdot v_\tau(S) + \frac{v_\tau(N) - v_\tau(\emptyset) - \sum_{i\in N}\left\{\sum_{S:i\in S} \frac{2^{|S|-1}}{3^{n-2}} \cdot v_\tau(S)\right\}}{n} \tag{16}$$

2.5.4 対数関数的に増加するカーネルに基づく AFA

(11) 式および (13) 式で定義されるカーネルは, それぞれ統計学における一様カーネル関数 (uniform kernel) と三角カーネル関数 (triangular kernel) に対応している. また, (15) 式のカーネルは, 凸型カーネル関数に対応している. 一方, 以下で定義される凹型カーネル関数は, エパネチニコフ・カーネル (Epanechnikov kernel) あるいはコサインカーネル (cosine kernel) に対応するものである.

$$\pi_x(S)^{CvK} = \frac{|S|(2n-|S|)}{(3n^2-n+2)\cdot 2^{n-4}.} \tag{17}$$

$\pi_x(S)^{CvK}$ に基づく AFA である $\Psi_{\tau,j}^{CvK}$ は, 以下の (18) 式の通りである:

$$\begin{aligned}\Psi_{\tau,j}^{CvK} = &\sum_{S:j\in S} \frac{|S|(2n-|S|)}{(3n^2-n+2)\cdot 2^{n-4}} \cdot v_\tau(S) \\ &+ \frac{v_\tau(N) - v_\tau(\emptyset) - \sum_{i\in N}\left\{\sum_{S:i\in S} \frac{|S|(2n-|S|)}{(3n^2-n+2)\cdot 2^{n-4}} \cdot v_\tau(S)\right\}}{n}.\end{aligned} \tag{18}$$

カーネルが (15) 式のように指数関数的に増加しているのであれば, 摂動サンプル z が分析対象となる観測値 x に近づくにしたがい, z に付与される重みは急激に増加していく. これは,「説明可能なモデル g が x 近傍で学習モデル f を近似しているかどうか」という点をより重視して最適な g を探索することを意味する. 一方, カーネルが (17) のように対数関数的に増加しているのであれば, x から離れた摂動サンプルにおける近似も比較的重視して g を探索することを意味する.

3　数値例を用いたSHAPとその代替的手法の比較分析

3.1　SHAPとその代替的手法のまとめ

本節および次節では, 前節までに示された様々なAFAを, 実際の数値例 (3節) および時系列金融・経済データ（4節）に適用して, 各AFAが特徴量に与える寄与度 (協力ゲームの文脈であれば利得ベクトル) の比較分析を行う. 分析対象となる6つのAFAをまとめると, 表1の通りである. 最初の3つが協力ゲーム理論の解概念に基づいたAFAであり, それぞれシャープレイ値 ($\Psi_{\tau,j}^{SHAP}$), 残余均等配分解 ($\Psi_{\tau,j}^{ES}$), 最小二乗プレ仁 ($\Psi_{\tau,j}^{PNucl}$) に対応している. 4番目から6番目は$|S|$に関して増加するカーネル関数から導出されたAFAであり, それぞれ$|S|$に関して線形に増加 ($\Psi_{\tau,j}^{LnK}$), 指数関数的に増加 ($\Psi_{\tau,j}^{ExK}$), 対数関数的に増加 ($\Psi_{\tau,j}^{CvK}$) するカーネル関数に基づいている.

表 **1:** 比較分析対象の **AFA** 一覧

記号	式	文献	特徴
$\Psi_{\tau,j}^{SHAP}$	(4)	[11]	シャープレイ値に基づくAFA
$\Psi_{\tau,j}^{ES}$	(5)	[3] [6]	残余均等配分解に基づくAFA
$\Psi_{\tau,j}^{PNucl}$	(12)	[6]	最小二乗プレ仁に基づくAFA
$\Psi_{\tau,j}^{LnK}$	(14)	[6]	線形増加するカーネル関数を持つAFA
$\Psi_{\tau,j}^{ExK}$	(16)	[6]	指数関数的に増加するカーネル関数を持つAFA
$\Psi_{\tau,j}^{CvK}$	(18)	[6]	対数関数的に増加するカーネル関数を持つAFA

3.2　数値例を用いた比較分析

まずは, 機械学習による予測という文脈を離れ, 各AFAを協力ゲーム理論における解とみなしたうえで, 単純な特性関数形ゲーム (N, v_τ) を用いて, どのような場合にAFA間の配分パターンの違いが明確になるか, また, その違いはどのように特徴づけられるかを考察する.

具体的には, 以下の 4 人のプレーヤー (特徴量が 4 つの場合に対応) から構成される特性関数形ゲーム ($N = \{1, 2, 3, 4\}$) を考える.

$$v_\tau(S) = \begin{cases} 50 & \text{if } S = \emptyset,\ |S| = 1,\ \text{または} |S| = 2 \\ 50 & \text{if } S = \{1, 3, 4\},\ \text{または} S = \{2, 3, 4\} \\ 90 & \text{if } S = \{1, 2, 3\},\ S = \{1, 2, 4\},\ \text{または} S = N. \end{cases}$$

このゲームの提携値 $v_\tau(S)$ の特徴をみると, 1 人提携, 2 人提携の提携値はプレイヤー間で完全に対称である. 一方, 3 人提携の提携値は非対称になっている. 具体的には, プレイヤー 1 および 2 については, 自らが提携に含まれる 3 つの場合のうち, 2 つの場合の提携値が 90, 1 つの場合の提携値が 50 になっている. 一方, プレイヤー 3 および 4 については, 2 つの場合で 50, 1 つの場合で 90 になっている. また, このことにより, 3 人提携から全体提携に変化する場合の各プレイヤーの貢献度をみると, プレイヤー 1 と 2 についてはそれぞれ 40 である一方, プレイヤー 3 と 4 については 0 となっている.

ここで, 全体提携値とは, AFA の文脈においては, 全ての特徴量が既知である場合の予測値 (すなわち, 分析対象である x における予測値) に対応することに留意されたい. そして, $|S|$ に関して増加するカーネル関数を持つ AFA は, x に近い摂動サンプルをより重視して説明可能なモデル g の近似度を評価するのであった. このことは, $|S|$ に関して増加するカーネル関数を持つ $\Psi_{\tau,j}^{LnK}$, $\Psi_{\tau,j}^{ExK}$ および $\Psi_{\tau,j}^{CvK}$ では, プレイヤーへの利得配分 (特徴量への貢献度の配分) にあたって, 規模の大きな提携の提携値をより強く勘案することを意味する. 一方, $|S|$ に関して独立なカーネル関数を持つ $\Psi_{\tau,j}^{PNucl}$ や, 増加関数の条件を満たさないカーネルを持つ $\Psi_{\tau,j}^{SHAP}$ や $\Psi_{\tau,j}^{ES}$ は, プレイヤーの数が少ない提携の提携値も相対的に重視することを意味する.

この点を踏まえて各 AFA が特徴量に与える貢献度をみると (図 6), まず, x との近接度を最も (すなわち指数関数的に) 重視する $\Psi_{\tau,j}^{ExK}$ が与える利得ベクトルは, 他の AFA と比べ, プレイヤー 1 と 2 に強く傾斜したものとなっている. これは, 3 人以上の提携値がプレイヤー間で非対称となっているという提携値の特徴が, より強く反映されているためである. $|S|$ に関して増加するカーネル関数を持つ $\Psi_{\tau,j}^{LnK}$ および $\Psi_{\tau,j}^{CvK}$ も同様の傾向がみられるが, 1 と 2 への利得配分の傾斜度は徐々に低下しており, 3 つの AFA が持つカーネル関数の増加パター

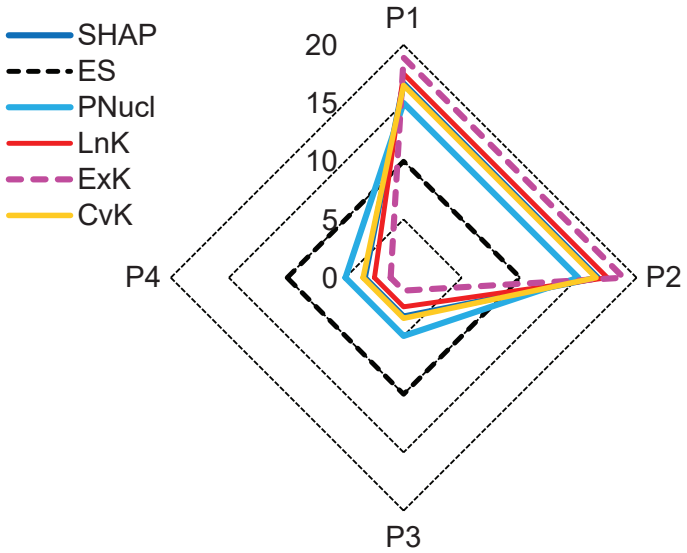
ンの違い (指数関数的か, 線形か, 対数関数的か) と整合的になっている. 一方, $|S|$ に関して独立なカーネル関数を持つ $\Psi_{\tau,j}^{PNucl}$ は, 規模の小さい提携も同程度に勘案する点を映じて, 3 および 4 にも相対的に多くの利得ベクトルを与えている. $\Psi_{\tau,j}^{ES}$ については, (5) 式から明らかなように, 全体提携と一人提携のみを考慮して利得ベクトルが決まることから, このゲームにおいては完全に均等な利得ベクトルとなっている.

図 6: 4 人ゲームの数値例における AFA

(1) 各 AFA が特徴量 (ここではプレイヤー) に与える値

		player 1	player 2	player 3	player 4	平均乖離度（%）
	SHAP	16.7	16.7	3.3	3.3	--
Proposed AFAs	*ES*	10.0	10.0	10.0	10.0	120.0
	PNucl	15.0	15.0	5.0	5.0	30.0
	LnK	17.5	17.5	2.5	2.5	15.0
	ExK	18.9	18.9	1.1	1.1	40.0
	CvK	16.5	16.5	3.5	3.5	2.6

(2) 上の表の値をグラフ化したもの



このように, 各 AFA が与える利得ベクトルは, そのカーネル関数の形状を明確に反映した

ものとなっている．また, これらの利得ベクトルと SHAP が与える利得ベクトルの平均的な乖離度をみると (図 6 上表の青いシャドー部分), 例えば $\Psi_{\tau,j}^{ExK}$ では 40%と, 相応に大きなものとなっている．このように, ゲームの構造によっては, AFA が与える利得ベクトルのパターンの違いはかなり大きくなる．このことは, 機械学習の予測モデルにおいて, SHAP のみによって特徴量の貢献度を判断することの危険性, 換言すれば, HIS 論文で提示された代替的な手法も用いて, 複数の手法・視点で総合的に判断することの重要性を示唆しているといえる．

そこで, 以下の 4 節では, 実際の金融・経済データに SHAP およびその代替的な手法を適用し, 各特徴量の貢献度のパターンに違いがみられるのか, 比較分析する．

4　SHAP とその代替的手法の金融・経済データへの適用

本節では, 実際の金融・経済データに対して様々な AFA, すなわち SHAP および HIS 論文で提示した複数の代替的手法を適用し, 各特徴量の予測貢献度に関する分解パターンが, 各手法の間でどの程度異なるのかを比較分析する．具体的には, (1) 1998 年以降のドル建て金価格, (2) 2001 年以降のわが国の有効求人倍率, の 2 つの時系列データを対象にする．比較分析の際に用いる手法は, (A) 時系列グラフによる視覚的な比較, (B) 特定の観測値のある特徴量において, 異なる AFA によって与えられた値の差の絶対値をベースにした比較, の 2 つである．

なお, 本分析の目標は, SHAP およびその代替的手法の分解パターンの違いを見ることであり, 機械学習モデルそのものの予測度や汎化性能（未知データへの適応力）を評価することではない．したがって, 以下では, 機械学習モデルは XGboost で固定し, かつ, 全てのデータを学習モデルとした In-sample の分析を行う.[*8]

[*8] したがって, 以下で示すグラフから分かるように, いわゆるオーバーフィッティングがみられるが, これも,「学習済みモデルを所与として, 異なる AFA 間の分解パターンの違いを比較する」という目的と照らし合わせると, ここでの論点とはならない．なお, XGBoost とは, 多数の決定木を順に構築し, 前のモデルの誤差を修正しながら予測精度を高めていく勾配ブースティング法をベースとした機械学習モデルである．

4.1　商品価格 (金価格)

Jabeur et al. [7] は, 金価格を 6 つの機械学習モデル (Linear regression, Neural networks, Random forest, Light gradient boosing machine, CatBoost algorithm, XGBoost algorithm) を用いて予測した後, これに SHAP を適用して比較分析を行い, XGBoost とそれに対する SHAP の適用が分析上有効であることを主張した. ここでは, Jabeur et al. [7] に基づき, 金価格の機械学習による予測モデルを構築し, それを 3.1 節の表 1 で示した, SHAP を含む 6 つの AFA で要因分解する.

図 7: 金価格の推移 (1998 - 2023 年)

具体的には, 分析対象は 1998 年 1 月から 2023 年 12 月までの, ドル建て金価格 (1 オンスあたり, 月次) データである (図表 7). 学習モデルは XGboost を用いる. Jabeur et al. [7] に沿って, 特徴量は以下の 6 つとし, それぞれ 1 か月のラグをとっている. また, 機械学習の標準的な手法に倣い, 各特徴量は平均 0, 分散 1 に標準化したものを学習データとして用いる.

- Silverprice: 銀価格 (ドル/オンス)
- Oilprice: 原油価格 (ドル/バレル)

- USD_EUR: ユーロ対ドルレート
- USD_CNY: 人民元対ドルレート
- CPI: 米国消費者物価指数 (指数, レベル)
- SP500: 米国株価 (SP500, ドル)

SHAP および HIS 論文で提示した複数の代替的手法にもとづく分解パターンの違いは, 図 8, 図 9 および図 10 に示されている. 赤い実線が実際の金価格の推移, 黒い実線が学習モデルによる予測値, そして棒グラフが各特徴量に割り当てられた AFA の値となっている. したがって, ある特定の月において, 各特徴量に対応する棒グラフを積み上げた高さは, その月の黒実線の高さと一致する. また, グラフはすべて 1998 年初からの累積変化幅ベースで示されている.

図 8: 金価格の **AFA** 分解 **(1)**

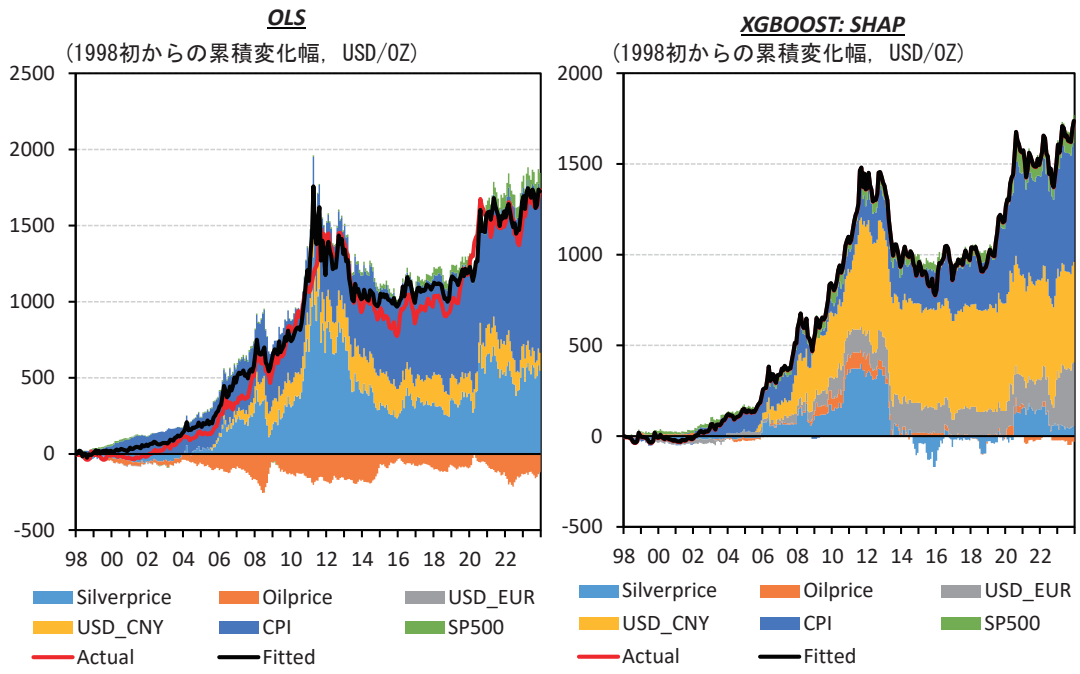
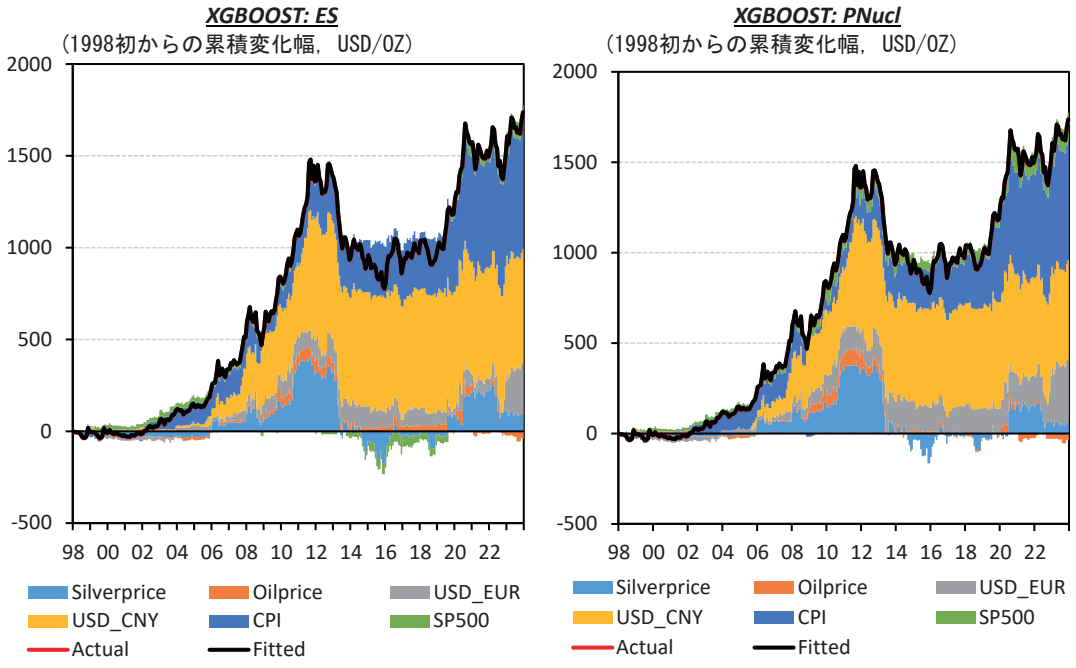


図 8 の左パネルは, 参考までに, XGBoost ではなく, 線形回帰 (OLS) モデルを学習モデルとして AFA 分解を行ったものである. 学習モデルが線形回帰モデルの場合には, 本分析で示したどの AFA を用いても, 分解パターンは同一になる (さらに, それは回帰モデルで推

定されたパラメータを用いた分解と一致する) ことが数学的に証明できる.[*9] 線形モデルと XGBoost の違いは本分析の主目的ではないのでこれ以上の詳述は行わないが, XGBoost と比べると, 回帰モデルによる予測は, 赤実線で示されている実績値と黒実線で示されている予測値の乖離が相対的に大きいこと, また, 線形性を仮定していることから, 各特徴量の寄与度の推移は, その特徴量自体の推移と比較的類似したパターンとなっていることがわかる.

図 8 の右パネルからが, 本分析の主な対象である SHAP および HIS 論文で提示された AFA による分解パターンを示している. 図 8 の右パネルが SHAP (表 1 における $\Psi_{\tau,j}^{SHAP}$), 図 9 は左パネルが残余均等配分解に基づく AFA ($\Psi_{\tau,j}^{ES}$), 右パネルが最小二乗プレ仁に基づく AFA ($\Psi_{\tau,j}^{PNucl}$) を用いた分解である. ここまでは協力ゲーム理論における既存の解概念をベースとした AFA である. 図 10 では, 特徴量の数について増加関数となっているカーネルに基づく AFA を用いた分解パターンを示している. 左パネルが指数関数的に増加するカーネル関数に基づく AFA (表 1 における $\Psi_{\tau,j}^{ExK}$), 右パネルが線形に増加するカーネル関数に基づく AFA ($\Psi_{\tau,j}^{LnK}$) である. なお, 対数関数的に増加するカーネル関数に基づく AFA については, グラフは示していないが, 後に示す表 2 において他の AFA との比較を行っている.

図 **9:** 金価格の **AFA** 分解 **(2)**



[*9] 証明については著者に問い合わせられたい. また, 今後別稿で示す予定である.

これらのグラフにおける分解パターンの違いを視覚的にみると, いくつかの点が分かる.

1 点目は, 全体としてみると, AFA の間の分解パターンは概ね類似しているということである. 例えば, どの AFA による分解パターンにおいても, 黄色の人民元対ドルレートや, 濃青の米国 CPI が金価格変動の要因となっている. 前者については, 2000 年から 2010 年代半ばにかけて, 人民元の切り上げが中国における金需要を高めたこと, また, 後者については, パンデミックをきっかけとした高インフレが, インフレヘッジとしての金の需要を高めたこと, を捉えたものであるといえる. こうした姿に AFA の間に大きな違いはない.

2 点目は, 協力ゲーム理論の解概念に基づく AFA を比較すると, シャープレイ値に基づく SHAP と最小二乗プレ仁に基づく AFA は, 子細に見てもほぼ同様の分解パターンとなっている一方, 残余均等配分解に基づく AFA は, これらとは相応に異なる分解パターンとなっている点である. SHAP と最小二乗プレ仁に基づく AFA は, カーネルの観点からは, U 字型か一定か, という違いであった. 少なくとも今回用いた金価格のデータでは, この程度のカーネルの違いであれば, 分解パターンに大きな違いをもたらすようなことはないことが明らかになった. 一方, 残余均等配分解に基づく AFA が, SHAP または最小二乗プレ仁に基づく AFA と比べて最も大きく異なる点は, 特徴量の数が中程度 (あるいは, 協力ゲームの言葉で言えば, 提携のサイズが中程度) である場合の予測値の情報を考慮していないということである. 実際のグラフを見ると, 例えば, 2014 年から 2019 年にかけて, 残余均等配分に基づく AFA では, 緑色で示されている株価 (SP500) が, 金価格に対してマイナスの寄与を示している一方, こうしたパターンは SHAP あるいは最小二乗プレ仁に基づく AFA では観察できない. また, 残余均等配分に基づく AFA は, 同時期の人民元対ドルレートのプラス寄与が, SHAP などと比べて相対的に大きくなっている.

3 点目は, 特徴量の数について増加関数となっているカーネルに基づく AFA について, 増加パターンの違い (指数関数的か (図 10 の左パネル) 線形か (同右パネル)) による違いは視覚的にはほどんど確認できない. また, これらの分解パターンと, SHAP や最小二乗プレ仁に基づく AFA も, 概ね同様の分解パターンを示している. したがって, この金価格の例においては, カーネル関数の違いが視覚的に見て大きな違いをもたらす, といった現象は (残余均等配分解に基づく AFA を除いては) みられず, むしろ SHAP およびその代替的な手法を用いた AFA による可視化分析の頑健性が示される結果となった.

図 10: 金価格の AFA 分解 (3)

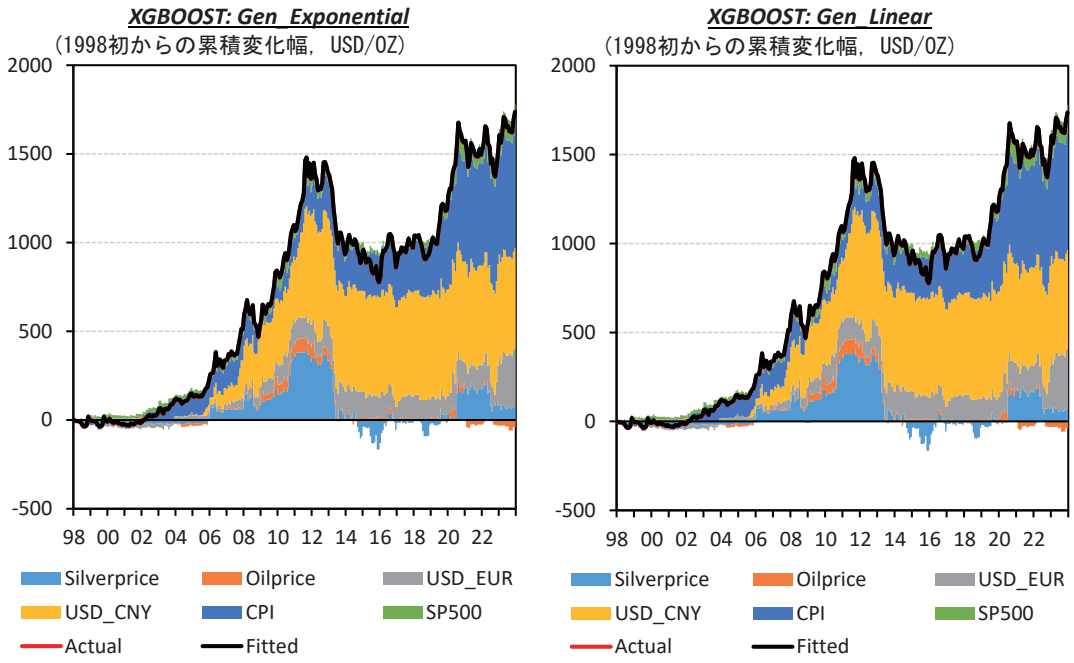


表 2: AFA 間の分解パターンの違い: 金価格

	SHAP	PNucl	ES	Exponential	Linear	Concave
SHAP	–	0.043	1.799	0.594	0.297	0.210
PNucl	0.106	–	1.808	0.598	0.298	0.206
ES	4.482	4.504	–	1.213	1.513	1.604
Exponential	1.479	1.490	3.022	–	0.300	0.392
Linear	0.740	0.743	3.768	0.747	–	0.092
Concave	0.523	0.514	3.996	0.977	0.229	–

注: 表 1 で示した 6 種類の AFA (Concave については, 上図においてはグラフの表示を省略している) のそれぞれの組み合わせについて, (1) 行列の右上半分で, 平均絶対差 (ある観測値におけるある特徴量について, 2 つの AFA の差の絶対値を計算し, それを全ての特徴量およびすべての観測値について平均したもの), (2) 行列の左下半分で, 平均絶対差 を予測値の標準偏差で標準化したもの, を表示したもの.

次に, 特定の観測値のある特徴量において, 異なる AFA によって与えられた値の差の絶対値をベースにした比較を行う. 表 2 で示された行列は, 右上の領域と左下の領域に分かれている. 右上の領域内のセルは, ある観測値におけるある特徴量について, 行に対応した AFA と列に対応した AFA の差の絶対値を計算し, それを全ての特徴量およびすべての観測値につい

て平均した値を示している. 左下の領域内のセルは, それを予測値の標準偏差で標準化した値を示している.

どちらの領域に着目しても, 以下の点が分かる. 1 点目は, グラフによる分析からも明らかになった通り, 残余均等配分 (ES) と他の AFA の差が相対的に大きいということである. 2 点目は, AFA 間の値の差は, カーネルの形状をある程度反映したものとなっているという点である. 例えば, 特徴量の数とは独立の, 一定のカーネル関数を持つ最小二乗プレ仁 (PNucl) をベースにした AFA を基準に, ES 以外の AFA との差をみると, 一定の範囲内で PNucl のカーネル関数と類似した形状となる, U 字型のカーネルを持つ SHAP との差がもっとも小さくなっている. さらに, 増加関数型のカーネルに基づく AFA と PNcul との違いをみると, 傾きが急激に高まる指数関数型のカーネルに基づく AFA (Exponential) がもっとも大きな差を示している一方, 傾きが徐々に緩やかになっていく Concave 型の AFA は PNcul との差が相対的に小さい. このように, 視覚的な観点からは判断できなかったものの, 異なる AFA が特徴量に与える値の差の絶対値に基づく分析結果は, カーネルの形状の違いを背景として, 異なる AFA 間の間で分解パターンの違いが生じうることを示唆している.

4.2 有効求人倍率

次に, わが国の有効求人倍率についての機械学習による予測モデルを構築し, それを前節同様, 3.1 節の表 1 にある 6 つの AFA で要因分解する.

対象となるデータは 2001 年 1 月から 2024 年 12 月までの有効求人倍率である (図表 11). 学習モデルは前節同様, XGboost を用いる. 特徴量については, SHAP を用いて米国の労働市場の分析を行った Buckmann and Joseph [2] 等に基づき, 以下の 8 つとし, それぞれ 1 期ラグをとる. また, 各特徴量は, 金価格のケースと同じく, 平均 0, 分散 1 に標準化したものを学習データとして用いる.

- Lag_Kyujin: 有効求人倍率 1 期ラグ
- 3MTB: 3 か月短期国債利回り (%)
- IIP: 鉱工業生産指数 (指数, レベル)
- NKY: 日経平均株価 (指数, レベル)

- LOAN: 国内銀行貸出 (前年比, %)
- CPI: 消費者物価 (指数, レベル)
- Oil: 原油価格 (ドル/バレル)
- M2: マネタリーベース (M2, 前年比)

図 **11:** 有効求人倍率の推移 **(1998 - 2024** 年**)**

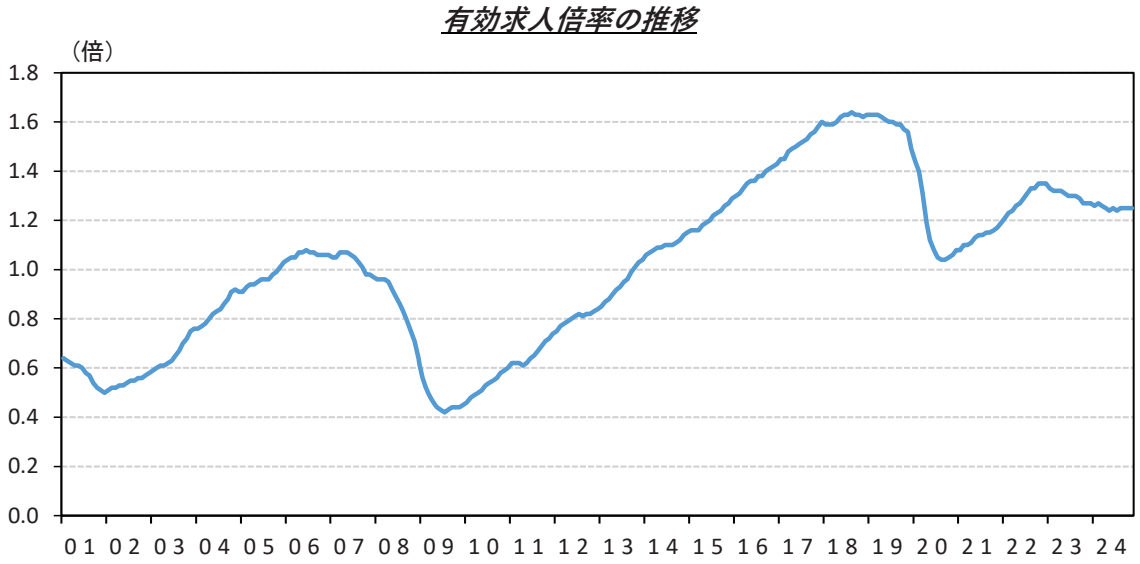


表 1 で挙げられた AFA (ただし前節同様 $\Psi_{\tau,j}^{CvK}$ は除く) にもとづく分解パターンの違いは, 図 12, 図 13 および図 14 に示されている. 赤実線が実際の有効求人倍率の推移, 黒実線が学習モデルによる予測値, そして棒グラフが各特徴量に割り当てられた AFA の値である. また, 分析の対象期間 (すなわちモデルの学習期間) は 2001 年以降であるが, ここでのグラフはすべて 2007 年以降に焦点をあて, 前年差ベースで表示している.

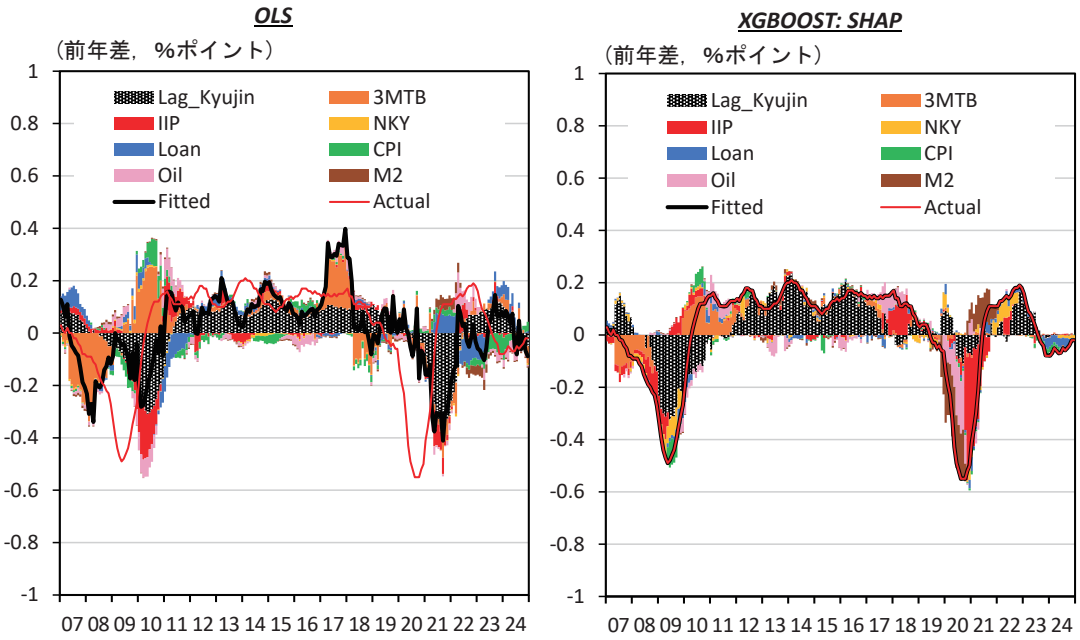
金価格のケースと同様, 図 12 の左パネルは, 参考として, 線形回帰モデルによる AFA 分解を示している. 図 12 の右パネルが SHAP ($\Psi_{\tau,j}^{SHAP}$), 図 13 は左パネルが残余均等配分解に基づく AFA ($\Psi_{\tau,j}^{ES}$), 右パネルが最小二乗プレ仁に基づく AFA ($\Psi_{\tau,j}^{PNucl}$) である. 図 14 は, 左パネルが指数関数的に増加するカーネル関数に基づく AFA ($\Psi_{\tau,j}^{ExK}$), 右パネルが線形に増加するカーネル関数に基づく AFA ($\Psi_{\tau,j}^{LnK}$) に基づく AFA である.

これらのグラフにおける分解パターンの違いを視覚的にみると, 以下の通り, 前節でみた金

価格におけるケースと同様の点が確認できる.

すなわち, 1 点目は, 全体としてみると, AFA の間の分解パターンは概ね類似している. 例えば, どの AFA による分解パターンにおいても, 2008 年の金融危機時における有効求人倍率の低下局面においては, 鉱工業生産 (IIP) の落ち込みが予測度の改善の主要因となっている一方, パンデミック時には IIP と原油価格の変動が主要因となっていたことが分かる. また, それぞれのショックからの回復局面では, 前者では短期金利, 後者は株価が求人倍率の回復に対する主要因となっていた点も, どの AFA における分解パターンにおいても観察される.

図 **12:** 有効求人倍率の **AFA** 分解 **(1)**



2 点目は, 協力ゲーム理論の解概念に基づく AFA を比較すると, 金価格のケースと同様, SHAP と最小二乗プレ仁に基づく AFA は, 視覚的には極めて類似した分解パターンとなっている一方, 残余均等配分解に基づく AFA は, これらとは相応に異なる分解パターンを示している. また, 3 点目についても, 金価格のケースと同様に, 特徴量の数について増加関数となっているカーネルに基づく AFA について, 増加パターンの違い (指数関数的か (図 14 の左パネル) 線形か (同右パネル)) による違いは視覚的にはほとんど確認できない. また, これらの分解パターンと, SHAP や最小二乗プレ仁に基づく AFA も, 概ね同様の分解パターンとなっている.

図 13: 有効求人倍率の AFA 分解 (2)

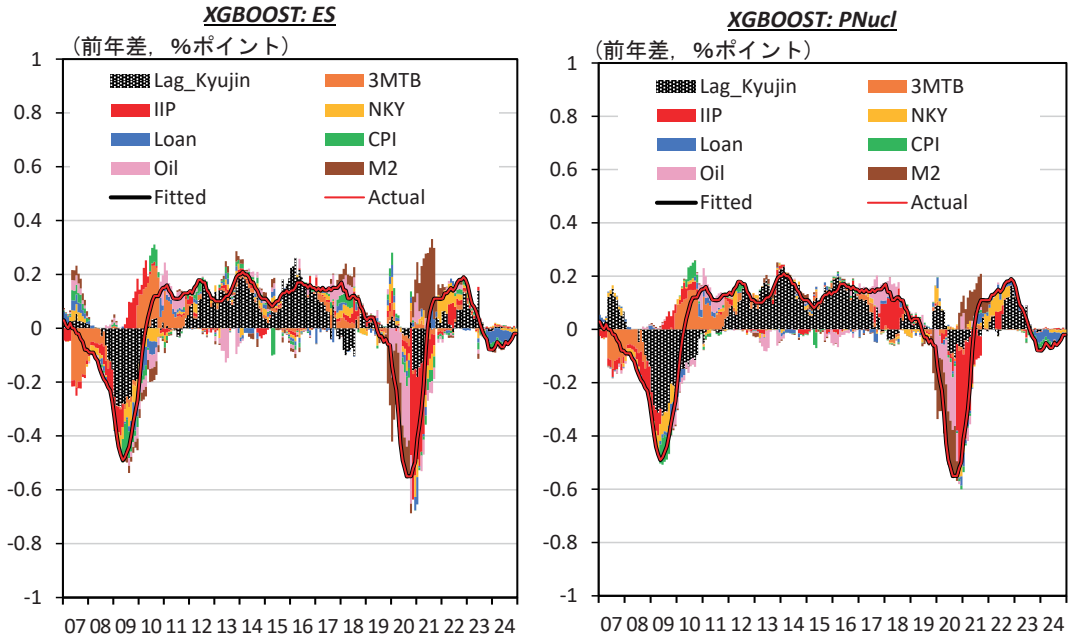
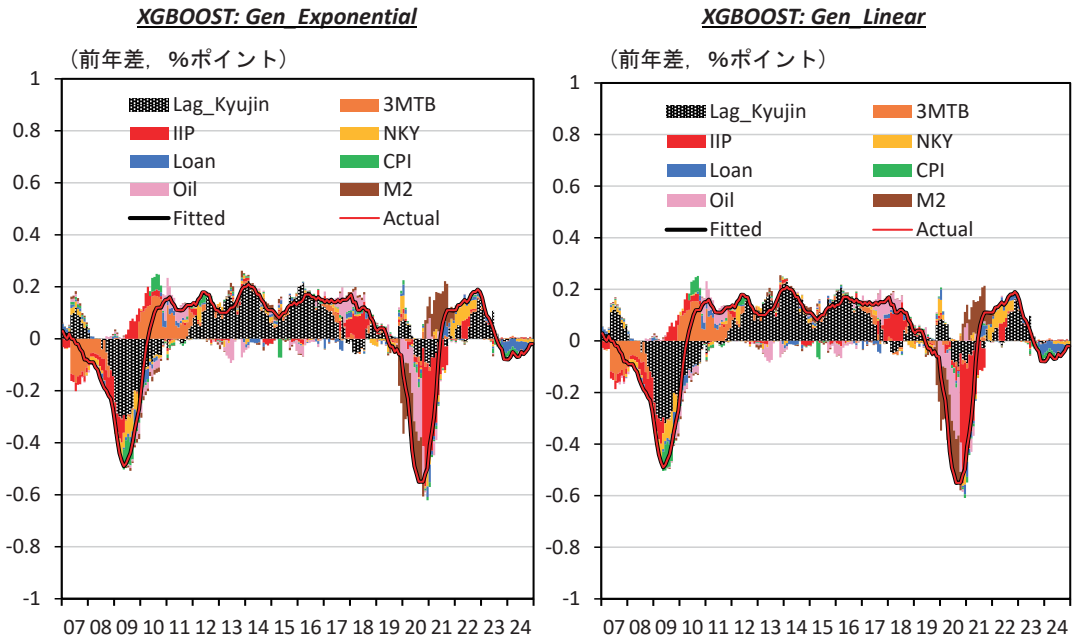


図 14: 有効求人倍率の AFA 分解 (3)



次に, 特定の観測値のある特徴量において, 異なる AFA によって与えられた値の差の絶対値をベースにした比較を行う. 表 3 は, 表 2 と同様に, 右上の領域内のセルには, 行に対応し

た AFA と列に対応した AFA の差の絶対値を計算し, それを全ての特徴量およびすべての観測値について平均した値が示されている. 左下の領域内のセルには, それを予測値の標準偏差で標準化した値が示されている.

ここでも, 金価格のケースと同様の特徴を観察することができる. すなわち, 1 点目として, 残余均等配分 (ES) と他の AFA の差が相対的に大きい. 2 点目に, AFA 間の値の差は, カーネルの形状をある程度反映したものとなっている. すなわち, 例えば左下半分の領域に着目すると, SHAP と PNcl の差は相対的に小さく, 次に Concave がこれらに類似している. 一方で, ES を除けば, 指数関数的にカーネルが増加していく Exponential は, SHAP や PNucl との差が相対的に大きい姿となっている.

表 **3: AFA** 間の分解パターンの違い**:** 有効求人倍率

	SHAP	PNucl	ES	Exponential	Linear	Concave
SHAP	–	0.001	0.014	0.005	0.002	0.002
PNucl	0.570	–	0.014	0.005	0.002	0.001
ES	8.373	8.509	–	0.010	0.013	0.013
Exponential	2.776	2.758	5.797	–	0.003	0.003
Linear	1.187	1.030	7.510	1.733	–	0.001
Concave	0.938	0.709	7.826	2.055	0.322	–

注: 6 種類の AFA (Concave については, 上図においてはグラフの表示を省略している) のそれぞれの組み合わせについて, (1) 行列の右上半分で, 平均絶対差 (ある観測値の, ある特徴量について, 2 つの AFA の差の絶対値を計算, それを全ての特徴量およびすべての観測値について平均したもの), (2) 行列の左下半分で, 平均絶対差 を予測値の標準偏差で標準化したもの, を表示.

5 まとめと結論

本項では, 高度に複雑な機械学習モデルを用いて計算された金融・経済データの予測値を, 人間が解釈可能な形に要因分解する AFA について概説し, 実際の金融・経済データへの適用可能性を検討した. 具体的には, まず, AFA の代表的な手法である SHAP と, Hiraki, Ishihara and Shino [6] で提示された SHAP の代替的な手法について, その特徴や理論的な背景を丁寧にレビューした. 次に, それらの手法を協力ゲームの数値例および実際のデータ (金価格および有効求人倍率) に適用して, 手法によってどの程度の要因分解のパターンの違いが生じるの

かを考察した. 協力ゲームの数値例における分析や, 実際のデータ分析において, 異なる AFA によって特定の特徴量に与えられた値の差の絶対値をベースにした比較からは, AFA の間で分解パターンに相応の違いがみられた. また, 観察された分解パターンの違いは, 各 AFA が持つカーネル関数の形状を反映したものであった. 一方で, グラフを用いた視覚的な比較においては, 残余均等配分解をベースにした AFA については他の AFA との違いが確認できた一方, 残余均等配分解をベースにした AFA 以外の AFA の間では, 視覚的にはっきりと確認できるほどの大きな違いは確認されなかった.

最後に, 本分析で得られたそのほかのインプリケーションを挙げつつ, それとの関連で今後更なる分析が有益であると考えられるいくつかのポイントを挙げて, 本稿を結ぶこととする.

(I) カーネル関数としての望ましい性質

本稿でも述べたように, カーネル関数を用いて機械学習モデルの解釈可能性を高める手法として, LIME（Local Interpretable Model-agnostic Explanations) がある. LIME を提示した Ribeiro et al. [13] においては, カーネル関数は,「実際の分析対象となる観測値に近い摂動サンプルほど, より大きな重みを与えて評価する」という考えに沿ったものであるべきとされている. 本稿で示した AFA のうち, この考え方をもっとも純粋に踏襲したものは指数関数的に増加するカーネル関数を持つ AFA (表 1 における $\Psi_{\tau,j}^{Ex^K}$) であるといえる. そして, 数値例や AFA の差の絶対値に着目した分析では, SHAP とこの AFA の間には, 相対的に大きな違いがみられた. 一方で, そうした違いが視覚的にも分かるほど明確なものになりうるかどうかは, 今回用いたデータから必ずしも明らかにはならなかった. 今後は, 様々なデータにこれらの AFA を適用することで, 特に U 字型のカーネル関数を持つ SHAP が, 不自然な分解パターンを示す場合があるかどうか, また, それは特にどのようなデータ特性において生じる蓋然性が高いのか, などについての分析を進める必要がある.

SHAP を用いた意思決定は, 医療や資金調達など, 社会経済活動の非常に重要な局面で急速に広がっていることから, その不安定性に対する理解を深めることや, 代替的な手法の開発に取り組むことは, 社会的にも非常に価値の高いテーマであるといえる.

(II) 計算コストへの対処

本稿でも述べたように, SHAP は, 計算コストが大きく, かつ特徴量の数が増えるほど指数関数的に増加することが知られている. SHAP を計算する際, Python などのソフトウェアにおいては, 近似計算用のパッケージがすでに利用可能であるが, それを用いたとしてもかなりの計算時間を要するケースが頻繁に生じうる.

本稿で示した AFA のうち, 残余均等配分解をベースにした AFA (表 1 における $\Psi_{\tau,j}^{ES}$) は, 計算コストが相対的に小さく, 特徴量の数が増えても計算コストの増加ペースは (SHAP のように指数関数的ではなく) 線形なものにとどまる. したがって, 特徴量の数が増えるほど, 両者の計算コストの差は大きくなる. そして, 協力ゲーム理論においては, 残余均等配分解の他にも, 計算コストを抑制することができる解概念が提示されている.[*10] 本稿では, 残余均等配分解による分解は, 他の AFA との違いが相対的に大きくなってしまったが, これは少ない計算コストで AFA を算出できることとコインの裏表の関係であるとも言える. 今後, 残余均等配分以外の解概念を用いたり, あるいはそれらを組み合わせる事によって, SHAP 等の分解パターンを近似しつつ, かつ大幅に計算コストを抑えることのできる AFA を開発・提示できれば, 学術的にも実務的にも大きな貢献となる.

(II) 様々なデータへの活用

本稿では, 実際の金融・経済への適用として, 時系列データを取り上げた. もっとも, 本稿で分析対象とした AFA は, 時系列データだけではなく, 原則, あらゆるタイプのデータに適用できる.

時系列データ以外への適用として興味深いものの 1 つとして, 株式のクロス・セクショナル・リターンを挙げておく. Gu et al. [5] は, 米国の約 30,000 銘柄の株式データを対象に, 決定木やニューラルネットワークといった機械学習モデルを用いて, 時系列だけでなく, クロスセクショナルなリスクプレミアムのモデル学習と予測を行った. 彼らは, 回帰ベースの従来手法と比較し, 機械学習モデルの予測精度が大幅に高まること, また, 実際の投資という観点からも, シャープレシオの改善などの経済的利得がもたらされることを示した. こうしたクロス

[*10] 例えば, Dragan et al. [4] や Kongo [17] が有益である.

セクショナルなデータを非線形の機械学習モデルで学習させ, それに基づく予測に対して本稿で用いた様々な AFA を適用し, 説明可能性を高めることによって, CAPM やマルチファクターモデルといった従来の資産価格モデルでは捉えることのできない関係性を把握する事ができる可能性がある.

参考文献

[1] K. Bluwstein, M. Buckmann, A. Joseph, S. Kapadia, and O. Şimşek. Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 145:103773, 2023.

[2] M. Buckmann and A. Joseph. An interpretable machine learning workflow with an application to economic forecasting. *International Journal of Central Banking*, 19–4:449–522, October 2023.

[3] C. Condevaux, S. Harispe, and S. Mussard. Fair and efficient alternatives to shapley-based attribution methods. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2023.

[4] I. Dragan, T. Driessen, and Y. Funaki. Collinearity between the shapley value and the egalitarian division rules for cooperative games. *OR SPEKTRUM*, 18:97–105, 1996.

[5] S. Gu, B. Kelly, and D. Xiu. Empirical asset pricing via machine learning. *Review of Financial Studies*, 33:2223–2273, 2020.

[6] K. Hiraki, S. Ishihara, and J. Shino. Alternative methods to shap derived from properties of kernels: A note on theoretical analysis. In *Proceedings of the International Conference on Big Data*, 2024.

[7] S. B. Jabeur, S. Mefteh-Wali, and J.-L. Viviani. Forecasting gold price with the xgboost algorithm and shap interaction values. *Annals of Operational Research*, 334:679–699, 2024.

[8] N. Jethani, M. Sudarshan, I. C. Covert, S.-I. Lee, and R. Ranganath. Fastshap:

Real-time shapley value estimation. In *International Conference on Learning Representations*, 2021.

[9] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and R. Acharya. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011 – 2022). *Computer Methods and Programs in Biomedicine*, 226:107161, 2022.

[10] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence volume*, 2:56–67, 2020.

[11] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, 2016.

[12] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, and S.-I. Lee. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2:749–760, 2018.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, New York, NY, USA, 2016. ACM.

[14] L. M. Ruiz, F. Valenciano, and J. M. Zarzuelo. The least square prenucleolus and the least square nucleolus. two values for tu games based on the excess vector. *International Journal of Game Theory*, 25:113–134, 1996.

[15] L. M. Ruiz, F. Valenciano, and J. M. Zarzuelo. The family of least square values for transferable utility games. *Games and Economic Behavior*, 24:109–130, 1998.

[16] L. S. Shapley. A value for n-person games. *Annals of Mathematics Studies*, 28:307–318, 1953.

[17] K. Takumi. Equal support from others for unproductive players: efficient and

linear values that satisfy the equal treatment and weak null player out properties for cooperative games. *Annals of Operations Research*, 338:973–989, 2024.

[18] 森いづみ、中村俊文、乗政喜彦. グローバルにみた感染症の家計等の行動への影響：機械学習によるアプローチ. **日銀レビュー**, 2021-J-5, 2021.

[19] 鷲見和昭. 通貨オプション市場における投資家センチメントの要因分析：機械学習アプローチ. **日本銀行ワーキングペーパー**, No.20-J-8, 2020.