【付録】ランダムフォレスト回帰と SHAP 値による分析手法の概要

今回のインサイト1[1]では、人的資本戦略と企業価値の関係を明らかにするため、機械学習のいち手法である「ランダムフォレスト回帰」と、その中身を可視化する「SHAP 値」を用いて分析を行った。

ランダムフォレスト回帰とは

ランダムフォレスト回帰は、複数の「決定木」という予測モデルを組み合わせて使う手法である。1本の決定木はデータに基づいて「こういう条件のとき、PBR（株価純資産倍率）や TFP（全要素生産性）はこのくらいになる」と予測するルールを作る。ランダムフォレストでは、そうした木をたくさん作って平均を取ることで、より安定して高精度な予測を可能にする。

SHAP 値とは

SHAP 値は、モデルが出した予測に対して「どの項目（特徴量）がどれだけ影響したか」を数値で示す手法である[2]。これはゲーム理論を応用した方法で、各要素の貢献度を公平に評価できるのが特徴である。SHAP 値を見ることで、「どの施策が企業価値を押し上げたのか、下げたのか」が一目で分かる。

SHAP 値（ϕ_i）は、モデルの予測結果を、各特徴量（あるいはその組み合わせ）がベースライン値（モデル全体の平均予測値など）からどれだけ予測値を押し上げたり下げたりしたかという「貢献度」として、ゲーム理論のシャープレイ値に基づいて公平に定量化するものである。これにより、個別施策の効果だけでなく、複数施策の組み合わせ効果も明確にできる。

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

- ϕ_i ： 特徴量 i の SHAP 値
- F： すべての特徴量の集合
- S： 特徴量 i を含まない特徴量の任意のサブセット
- $f_x(S)$ ： 特徴量のサブセット S のみが与えられた場合のモデルの予測値
- $|S|$ ： サブセット S の特徴量の数
- $|F|$ ： 全特徴量の数

[1] MRI オピニオン　2025 年 9 月号　インサイト1：企業価値を高める人的資本戦略とは？

[2] SHAP 値の概念を提案した基礎的な文献としては、Lundberg and Lee (2017)が、マクロ経済分野に応用した研究には Buckmann and Joseph (2023)などがある。

Copyright (C) Mitsubishi Research Institute, Inc.

この計算により、SHAP 値は以下の特徴を持つ。

- 公平な貢献度 ： 異なる特徴量間で、予測への貢献度を公平に比較できる。
- 加法性 ： すべての特徴量の SHAP 値の合計は、特定の予測値とベースライン値の差に等しくなる。

$$f(x) = E[f(x)] + \sum_{i=1}^{M} \phi_i$$

ここで、$f(x)$はモデルの予測値、$E[f(x)]$ はベースライン値（例えば、全データの平均予測値）、ϕ_i は各特徴量の SHAP 値、M は特徴量の総数である。

- 方向性 ：SHAP 値の符号（プラスかマイナスか）は、その特徴量が予測値をベースラインから「押し上げた」のか「押し下げた」のかを示す。

今回の分析での活用

今回の分析では、人的資本に関する6つの取り組み要素のほか、それらを組み合わせた「交差項」も評価対象とした。SHAP 値を使うことで、「複数の施策を組み合わせて実施したときに企業価値がどう変わるか」という相乗効果（シナジー）まで可視化することができる。実際に、単独の施策よりも、戦略的に束ねて取り組んだ場合に大きな効果が出ることが確認された。

このように、ランダムフォレストと SHAP 値の組み合わせは、人的資本と企業価値の関係を、従来の統計手法では捉えきれなかった形で「見える化」する強力な手段となる。

Copyright (C) Mitsubishi Research Institute, Inc.