

## 機械学習を用いた不正会計の検出と分析

田嶋 俊紀(21X4037) 平田 円(21X4121) 指導教員 劉 慶豊

## 1. はじめに

会計不正は、財務諸表を意図的に操作して実際の財務状況を歪める行為であり、業績向上へのプレッシャーや競争圧力などが動機として考えられる。不正企業は業績を良く見せる操作が大規模な経済的損失や規制強化を招いた。

先行研究において、会計不正の検知モデルにロジスティック回帰が用いられる場合は殆どである。Dechow(2011)[1]のFスコアを基にしたモデルや、発生費用関連のHealy(1985)[2]やJonesモデル[3]代表的なモデルである。資産指標を通じた利益操作の影響についてはBarton and Simko(2002)[4]で議論されている。不正会計の検出は近年においても世界的に盛んな研究分野であり、多数の研究が行われている。

近年不正会計の増加や手段の巧妙化が問題となっており、より効率的な監査が求められている。そこで本研究は機械学習を用いて不正検出における財務諸表項目の重要度を分析して監査効率化を図る。具体的には財務諸表のデータを分析することにより、不正の予測に対して重要度が高い特徴量を検出する。さらに、会計学の視点からの分析も行い、予測の上での重要度が高かった特徴量と不正会計との関係を解明することを目的とする。改めて不正会計が起きてしまう要因は何かに関して分析、考察することで、会計監査や会計基準の改善に寄与することを目指す。

## 2. データについて

本研究では、2020~2024年に提出された上場企業の財務諸表データを使用する。対象とする。特に財務諸表は、資産、負債、純資産を示し、財務状況を把握する貸借対照表、収益から費用を引くことで収益を示し、経営成果を数値化する損益計算書、発生に基づき現金の流れを記録し、営業活動、投資活動、財務活動に分けて資金繰りや財務健全性を評価するキャッシュフロー計算書をこの研究で扱う。

データはデジタル庁の公開データセット(<https://catalog.registries.digital.go.jp/rc/dataset/>)と金融庁の「課徴金納付命令勧告」に書かれている企業について

EDINET(<https://disclosure.edinet->

[fsa.go.jp/WEEK0010.aspx](https://fsa.go.jp/WEEK0010.aspx))から取得した。約12,000社から約4,000社のデータを抽出する。これに加え不正が発覚された56社のデータを用いる。

## 3. 分析手法

本研究ではデータ分析にランダムフォレストを用いる。

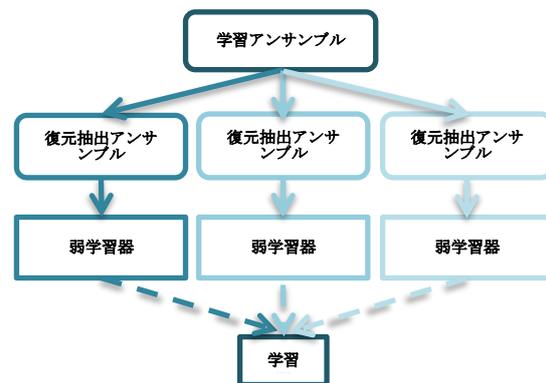


図1. ランダムフォレストにおける弱学習器の結合

## 3.1. ランダムフォレストの構造と特徴

ランダムフォレストは、複数の決定木を使って予測を行うアンサンブル学習の手法である。その構成要素である決定木は木構造を持ち、分類や回帰に使われる。ランダムフォレストでは「バギング」を用いて学習データをランダムに分割し、複数の弱学習器を作成して、最終的に、それらの弱学習器による予測結果を統合する。このようなアプローチでは並列処理が可能となるため、計算効率が高い。また、精度向上のため、決定木同士の相関を小さくすることが重要である。ランダムフォレストを用いた分析の中で、特徴量重要度を計算することで、各特徴量が予測にどれだけ貢献したかを確認できる。

## 4. 実装

## 4.1. データの前処理

第4章でダウンロードしたHTMLファイルをスクレイピングしてデータフレーム化し、CSVファイルとして保存した。不正を示す「ラベル」列を追加し、不正データを1、正常データを0とした。財務諸表の項目を特徴量、不正判定をターゲットとして、データを

70%のトレーニングセットと30%のテストセットに分割した。

#### 4.2. モデル構築と学習

ランダムフォレスト分類器を使用し、決定木の数を100に設定した。過学習を防ぐため複数の木を平均化し、木の深さは制限せずトレーニングデータで学習させた。

#### 4.3. モデル評価

テストデータで予測を行い、正解データと比較して正解率で予測精度を評価した。

### 5. 分析結果

モデルの学習後、特徴量の重要度の一種である Feature Importance (FI)を算出した。FIは特徴量の予測への寄与度を示し、合計が1となるように正規化されている。各年の結果の一例として2023年の結果を表1に示す。

表 1. 2023 年に関する特徴量重要度（上位を抜粋）

| 特徴量                        | 重要度   |
|----------------------------|-------|
| 有形固定資産・無形固定資産の増加額          | 0.041 |
| 1株当たり当期純利益又は当期純損失(Δ)、経営指標等 | 0.025 |
| 為替換算調整勘定(税引後)、その他の包括利益     | 0.018 |
| 経常利益又は経常損失(Δ)、経営指標等        | 0.016 |
| 受取配当金、営業外収益                | 0.015 |

2020年の場合、キャッシュフローや連結会計項目の多くが重要度の高い項目となっている。

また、資金の動きを反映する項目が目立つ。2021年はキャッシュフロー計算書や損益計算書において貸倒引当金に関する項目の重要度が高い。また債権や固定資産の売却・取得が目立つ。2022年は為替や外貨との換算について重要度の高い項目が頻出した。また資金の動きを反映する項目も多く見られた。2023年の場合、資産に関する項目が重要度の高い項目になった。また利益に関する項目も多く見られる。2024年は「福利費」や「租税公課」が上位で、営業外費用や特別利益が重要な項目になった。

各年に渡って、重要度のばらつきが見られたが、全体としては、キャッシュフロー計算書や損益計算書項目が多く、特に、「有形固定資

産及び無形固定資産の増加額」「長期借入金の返済」「1株当たり当期純利益」が共通して重要であった。

### 6. 考察

以上の結果を踏まえてまず、資産の過剰評価や減損回避を目的に、不良企業が企業規模を誇張する操作を行おうとする可能性があるため、「有形・無形固定資産の増加額」は不正企業を検出する上で重要な指標であると考えられる。また、「1株当たり当期純利益(EPS)」は利益水増しや費用過少計上による利益誇張、「長期借入金返済額」はキャッシュフロー操作による財務健全性誇張を検出するために重要な指標であると考えられる。損益計算書は利益や費用の操作余地が大きい一方、キャッシュフロー計算書は現金の流れを示し、損益計算書との矛盾や異常取引が見えやすく、不正検出上、重要な項目として重視されるべきである。

### 7. まとめ

本研究では、財務諸表の数値データを用いてランダムフォレストによる重要度分析を行い、会計不正検知において重要である特徴量を特定した。監査業務を行う上で重視すべき財務指標を明らかにしたことで、監査業務の有効化と効率化へ大きく貢献すると期待できる。本研究は財務諸表以外の情報を考慮していない点が限界であり、非数値データを活用して、モデルをさらに改善していく課題が残る。

### 参考文献

- [1] Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary accounting research*, 28(1), 17-82.
- [2] Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. *Journal of accounting and economics*, 7(1-3), 85-107.
- [3] Jones, J. J. (1991). Earnings management during import relief investigations. *Journal of accounting research*, 29(2), 193-228.
- [4] Barton, J., & Simko, P. J. (2002). The balance sheet as an earnings management constraint. *The accounting review*, 77(s-1), 1-27.