

# 機械学習を用いたポートフォリオスクリーニング 手法の有効性に関する研究

## Study on equity portfolio screening through machine learning

韓 原草<sup>1</sup> 菅 愛子<sup>1</sup> 高橋 大志<sup>1</sup>

Yuancao Han<sup>1</sup>, Aiko Suge<sup>1</sup>, Hiroshi Takahashi<sup>1</sup>

<sup>1</sup>慶應義塾大学大学院経営管理研究科

<sup>1</sup>Graduate School of Business Administration, Keio University

**Abstract:** Stock prices fluctuate frequently reflecting new information such as released financial statements and news articles. For equity fund managers, stock performance is one of the main concerns. In this article, we attempt some trials to develop several portfolio screening methods which remove poor performance equities from equity portfolio using various kinds of machine learning.

### 1. はじめに

資産運用の重要性は高まっている。資産運用ビジネスにおいて、利潤を追求する株主や投資家は、独自に組成したポートフォリオへの投資を通じ、収益の獲得を試みている。近年人工知能のアプローチへの関心が高まっており、金融分野においてもその応用が幅広く議論されている。例えば、国内の大手証券会社の人工知能を活用した取り組みや[1]、景況感の指数化などの試みなどが報じられている[2]。

本研究では、株式ポートフォリオのスクリーニングに機械学習を応用することを試みる。具体的には、TOPIX (Tokyo Stock Price Index、東証市場第一部に上場する内国普通株式全銘柄を対象とする株価指数)に含まれる株式の中から保有しない銘柄を機械学習にて分類し除外する。本研究では、スクリーニングを実施した後、ポートフォリオの収益率が、どれだけ影響を受けるかについても分析を試みる。

### 2. 関連研究

財務関連指標を用い倒産する企業に関する分析を行った分析は数多いが、最も広く知られたモデルの一つに Altman [3]が挙げられる。流動性、収益性、レバレッジ、返済能力、回転率などを基準として、判別に必要な5つの財務指標を選択し、モデルを構築している。また、Ohlson [4]は、マクロ経済の指標を関連する指標の一つとして使用しており、Tyler Shumway [5]は、個別の株式のリスクを表す $\beta$ を変数の一つとして使用している。分析手法については、Ramakrishnan, Mizaei, Bekry [6]の研究では、単一の分

類器とアンサンブルの両方を使用し、それぞれの結果について比較分析を行っている。

### 3. モデル

本分析では、複数のアルゴリズムを用い AUC 値が最も高いモデルを選択し、銘柄の除外を行った。表1は、本分析で採用したアルゴリズムのリストを示したものである。また、比較対象のモデルとして最も広く用いられている分類アルゴリズムの一つである決定木を採用した。

表1.本分析で採用したアルゴリズム (抜粋)

アルゴリズム
決定木
ブースティング
バギング
KNN
SVM
ロジスティック回帰
線形判別
二次判別

図1はポートフォリオスクリーニングルールを示したものである。本分析では、3つのルール(ルールA-C)の有効性について検証を行った。ルールAとBにおいて、一つ目の教師変数はTOPIXすべての銘柄の中で、株価変動率が下位150の銘柄を-1と設定し、それ以外すべての銘柄を+1とした。また、二つ目は下位300の銘柄を-1とし、それ以外の銘柄を+1とした。三つ目は下位500の銘柄を-1であり、それ以外

は+1である。ルールCでは上位150、300、500をそれぞれ+1とつけ、下位150、300、500をそれぞれ-1とした。

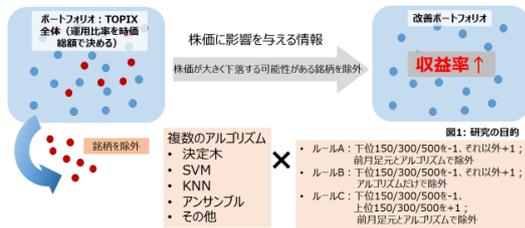


図1：ポートフォリオスクリーニングルール評価システムの構築

除外基準に直近の株価変動も考慮した。具体的には、ルールAとCでは足元の株価変動が下位にあるもので、翌月のアルゴリズムの推定値が下位(-1)のものを除外した。ルールBではモデルの推定値が下位(-1)の銘柄を除外した。

#### 4. データ

本分析では、Thomson Reuters Datastream から、日本の東京証券取引所で一部上場している企業の財務データと市場データを取得した。分析対象期間は2013年1月から12月であり、月次データを用いた。説明変数として、財務および市場関連データ、各企業のニュース記事数を用いた (表1)。

表1. 学習に用いた変数リスト(抜粋)

マーケットデータ		財務データ	
株価	売呼値	買掛金	投資利益率
市場価値	買呼値	売上債権回転期間	一株当たり売上高
配当利回り	最高値	簿価	負債/総資産
株価収益率	最低値	流動比率	ワーキングキャピタル
株価指数	return index as paid	配当性向	一株当たり簿価
12カ月前のEPS予想	トータルリターンインデックス	EBIT	資本的経費
過去12カ月のEPS	出来高	EBITDA	現金
一株当たり配当	未調整始値	企業価値	現金配当
一株当たり当期純利益	未調整株価	現金及び現金同等物変化	一株当たりキャッシュフロー
市場価値	未調整出来高	時価総額	キャッシュフロー/売上
		売上原価	発行済み株式数
		当座比率	ROE

#### 5. 分析結果

##### 5.1 除外前後の平均収益率比較

図3は、スクリーニング後のポートフォリオの超過収益を示したものである。横軸は、モデルの条件、縦軸は超過収益率を示している。ルールAでは除外後の平均収益率は教師変数が下位150と下位500の

場合、ポートフォリオのパフォーマンスが向上したが、下位300の場合は悪化している。ルールBとCはすべての教師変数の設定で悪化している。本条件においては、一部の条件で良好な結果が得られているものの、安定的な結果は得られていない。

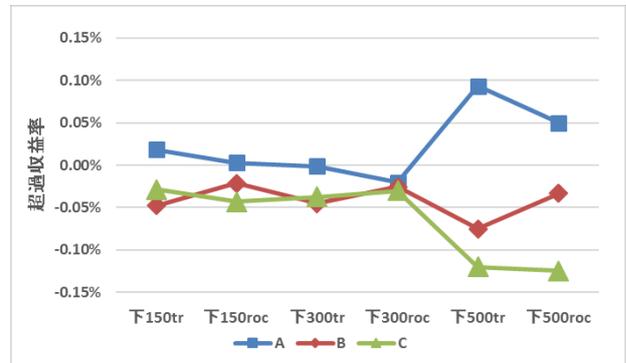


図3.スクリーニング後ポートフォリオ超過収益率

#### 6. まとめ・今後の課題

本研究では、機械学習手法を用いた株式ポートフォリオのスクリーニング手法の構築を試みた。分析の結果、本条件では、必ずしも安定的な結果は得ることができなかった<sup>1</sup>。今後、ティックデータやニュースなど即時性の高い説明変数を採用することで、精度の向上を試みる予定である。

#### 参考文献

- [1] 「大和証券と大和総研、人工知能 (AI) による株価予測モデルを用いた銘柄情報の提供を開始」, [http://www.nikkei.com/article/DGXLRSPP444457\\_Z00C17A5000000/](http://www.nikkei.com/article/DGXLRSPP444457_Z00C17A5000000/), (2017年5月9日)
- [2] 山本 裕樹: 「ディープラーニングの経済・金融テキストへの応用」, JAFEE 高頻度データ・ビッグデータ活用法研究会, (2017年4月11日)
- [3] Altman, E. I.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23 (4) , 589, (1968)
- [4] Ohlson, J. A.: Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18 (1) , 109, (1980)
- [5] Shumway, T: Forecasting Bankruptcy More Accurately: A Simple Hazard Model. SSRN Electronic Journal, (2001)
- [6] Ramakrishnan, S., Mirzaei, M., & Bekri, M: Corporate Default Prediction with AdaBoost and Bagging Classifiers. Jurnal Teknologi, 73 (2) , (2015) .

<sup>1</sup> 市場の効率性の議論と整合的に捉えることができるかもしれない。詳細な分析は今後の課題である。