

# 機械学習を用いた

## ポートフォリオスクリーニング手法に関する研究

### Developing equity portfolio screening method 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 and Administration, Keio University

**Abstract:** Stock prices fluctuate frequently reflecting new information such as released financial statements and news articles. For equity fund manager, stock performance is one of the main concerns. In this article, we develop a portfolio screening method which removes poor performance equities from equity portfolio.

## 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、企業の将来性を予測し、独自に組成したポートフォリオから収益の獲得を試みている。株式価格は、企業価値を示す指標であり、その企業の状況や将来の成長性を表している。株式価格は、将来の企業業績にも大きく影響されることから、決算短信などを通じて開示される事業予測の数値などにも大きく影響される。また、ネガティブな要素にも市場は過敏に反応し、例えば企業が社会に対して何らかの不祥事を起こした場合、媒体を通じて発信されたニュースや情報はその企業の株価に大きな影響を与えることが知られている。

また、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が人工知能を活用し、株価予測モデルの導入と国内株式に関する情報のサービス提供との新聞記事や[1]、人工知能モデルを用いて景況感の指数化を行い、指数の提供を開始したなどの記事が報じられている [2]。

## 2. 目的

株式価格に影響を与える要因は数多く存在するが、決算短信や企業に関するニュースなどは主要な要因の一つに挙げられる。もし、このような情報に基づき、将来株価が大きく下落する可能性がある銘柄をポートフォリオから除外し、ポートフォリオ収益率

をあげることができれば、投資家にとって非常に有用である。

近年、機械学習アルゴリズムの発達に伴い、これまで人が定性的に行ってきた投資判断をモデル化できないかとの議論が関心を集めている。このようなモデル化は、ポートフォリオの迅速な構築に大きく貢献することが期待される。これらを背景として、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。更に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。

## 3. 先行研究

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例は多いが、最も広く知られたモデルの一つに Altman[3]が挙げられる。Altman は、貸借対照表と損益計算書の中から 22 個の財務関連変数を分析対象とし、流動性、収益性、レバレッジ、償還能力、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、Suresh Ramakrishnan,

Maryam Mizaei, Mahmoud 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。更に、近年、ディープラーニングの手法を用いた研究なども報告されている[7]。機械学習に関する研究は、これまで数多く行われているが、情報技術の進展を背景とし、近年、それら技術への関心は高まっている[8][9]。

本研究では、財務関連指標、機械学習の手法を用い、株式ポートフォリオのスクリーニングモデルの構築を試みる。

#### 4. 使用データ

本研究で扱う分析データは S&P Capital IQ から取得している。今回のポートフォリオ分析対象企業には 2014 年の第 3 四半期 (7-9 月) の期初に東京証券取引所一部上場していた企業、おおよそ 1800 社程度を用いた。企業データには貸借対照表、損益計算書、キャッシュフロー計算書といった三種類の財務諸表と市場データを抽出した。それら学習に用いる 20 あまりの変数については、表 1 のように、貸借対照表の指標、損益計算書、市場データ、比率指標、という 4 種類に分類することができる。これらデータで欠損値が多いなどの企業について、機械学習のモデル化の過程で除外しており、おおよそ 1700 社のデータを用いた。

表 1: 学習に用いた変数のリスト (抜粋)

貸借対照表	損益計算書	市場データ	比率
ワーキング キャピタル	売上高	時価総額	ROA
現金	EBIT	一株当たり配当	ROE
総資産	営業利益	一株あたり資産	売上成長率

また、ポートフォリオ収益率の分析については、日次の株式データを Thomson Reuters Datastream から取得した。

#### 5. 分析方法・結果

ポートフォリオスクリーニングに用いる機械学習法は以下のように選択した。まず、分析の基準、比較対象として、分類問題において広く用いられる決定木による分類を実施した。更に、サポートベクターマシン (SVM)・k 近傍法・ブースティング・バギングといった分類法を複数評価し、もっとも予測精度が高いモデルを各期で採用した。予測精度の指標

には、本分析では、ROC 曲線 (Receiver Operating Characteristic curve、受信者動作特性曲線) に基づく AUC (Area Under the Curve) を用いた。これらの手順を経て選択された分類器が、表 2 にまとめられている AUC が 0.62 であった (1) Boosting と (2) RUSBoost[10]である。なお、ポートフォリオから除外する銘柄を示す教師データとして、本分析では、直近 3 ヶ月の株式の超過収益率を採用した。具体的には、スクリーニング対象四半期の 1 期前の株価変動率 (ここでは第 2 四半期、4-6 月の変動率) が下位 1 割に入る企業を、除外銘柄の negative 教師データとして設定した。またモデルの変数には、さら前の財務・会計データを用いることとした。

表 2: 学習モデルと AUC

2014 Q3	分類器	AUC
共通モデル	決定木	0.53
AUC の高いモデル (1)	Boosting	0.62
AUC の高いモデル (2)	RUSBoost*	0.62

これらのモデル・データセットを用いて、negative 判定された銘柄の数は分類器によって異なることが確認できた。未だいずれの手法も多く偽陽性や偽陰性を含んでおり、モデルパラメーター最適化が今後の検討課題である。また、スクリーニング結果が、ポートフォリオ収益率にどのような影響を与えるのといった詳細な分析は追って実施する。

#### 6. まとめ・今後の課題

本研究は、機械学習手法を用い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ものである。学習変数に様々な企業データを用い、四半期ごとのスクリーニングを計画した。現在、用いる変数・分類器の種類・モデル精度によって、除外候補銘柄の数や種類が異なってくることを確認している。今後、これらの結果を用い、ポートフォリオの収益率がどのように変化するか検証する予定である。

また、現在のモデル化に用いる変数には財務・会計データを多く用いている。これらの変数は、ティックデータやニュースの更新速度と比べれば比較的即時性に劣る変数のため、ポートフォリオのパフォーマンスに与える影響が限定的であることが考えられる。今後、ティックデータやニュースなども分析対象とし、モデルの精度の向上を検討している。

#### 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/DGXLRSP444457\\_Z00C17A5000000/](http://www.nikkei.com/article/DGXLRSP444457_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)
- [7] 池田欽一, 林田実: 「ディープラーニングの株価予測への応用」, 北九州市立大学『商経論集』第52巻第1・2・3・4合併号, 13-26, (2017)
- [8] Bishop, C.M.: *Pattern Recognition and Machine Learning*, Springer-Verlag New York, (2006)
- [9] Hastie, T., Tibshirani, R., & Friedman, J.: *The Elements of Statistical Learning*, Springer-Verlag New York, (2009)
- [10] C. Seiffert, T.M. Khoshgoftaar, J. Van Hulse and A. Napolitano, "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance, *IEEE Transaction on Systems, Man and Cybernetics-Part A: Systems and Human*, Vol.40(1), January 2010.