

# Fusion of Sentiment and Asset Price Predictions for Portfolio Optimization

キム ジビン

**概要:** 株ポートフォリオの最適化を紹介する。近年、感情分析を利用した株の価格予測は金融コミュニティでは既に注目されている。だが、ポートフォリオ選択方法に関する研究は未だに少ない。本論文では、感情分析を利用したポートフォリオ選択方法の改善を目的とする。具体的には、Semantic Attention モデルを利用して資産に対する感情を予測する。そして、価格予測と mean-variance 戦略のため LSTM モデルを活用して最適なポートフォリオを選ぶ。我々の感情ポートフォリオ戦略は非感情ポートフォリオより収入率を改善した。だが、資産運用の安全性の面では、旧来のポートフォリオより優れる結果を出すことができなかった。

## 1. はじめに

ポートフォリオ最適化は本質的に複雑な金融計算問題である。ポートフォリオ選択とは、最も良い富の配分を探すことである。研究者らは様々な方法でこの問題を解決しようとしている。その中で最も有名な研究は、Paskaramoorthy らの研究 [1] である。近年では、投資家の感情を特徴量として株価を予測するモデルが人気になっている。Ferreira らの研究 [2] のように価格予測や感情分析、ポートフォリオ最適化の技術を融合することで良い結果を出している。今回の論文では、感情認識ポートフォリオ選択の強化を調査する。はじめに、今回の実験で使用する感情分析モデル VADER の精度を確認する。この結果から VADER モデルを改善した感情モデル Semantic Attention model を使う動機となる。感情認識モデルは上昇トレンドと下降トレンドの市場で従来のポートフォリオ配分戦略を上回る。ただし、この戦略は従来のアプローチよりも堅牢性が低く、リスク率が増加した。研究の結果では、2つのポートフォリオ最適化を比較する。1つは従来の mean-variance ポートフォリオ最適化によるもので、もう1つは資本の最適な配分を予測する LSTM フレームワークによるものである。この調査は、感情を意識したポートフォリオ戦略が、従来のポートフォリオ選択方法よりも高いリターンを生み出すことを示している。そして、実際に価格と投資家の感情が相関していることを詳しく説明する。また、これらの結果から感情分析ツールには偏りがあることも示す。

## 2. 提案手法

### 2.1 価格データ

ポートフォリオは Dow Jones Industrial Average (DJIA) から5つの企業で構成。価格データは Yahoo Finance の 2001 年 1 月 1 日から 2018 年 12 月 31 日までの終値調整後の価格を使用する。70 % の訓練データ、20 % のテストデータ、10 % の検証データに分割される。訓練データは 2001 年 1 月 1 日から 2013 年 12 月 11 日まで、検証データは 2013 年 12 月 12 日から 2015 年 5 月 20 日、テストデータは 2015 年 5 月 21 日から 2018 年 12 月 31 日までとする。

### 2.2 テキストデータ

感情データは Google News と Twitter から取得する。データの収集方法は、スクレイピングを使用する。具体的には、ニュース記事は Selenium WebDriver を使用して 5 社に関するニュース記事をスクレイピングする。ツイートデータは、Snsrape を使用してスクレイピングする。感情データは、価格データと同様の期間で訓練データ、検証データ、テストデータに分割される。また、提案手法では、7日間感情データを一つのウィンドウに設定する。

### 2.3 ラベリング

本実験のデータには感情ラベルが付けられていない。そのため、ニュース記事やツイートデータの内容では感情を含んでいるかどうか認識することができない。このような場合、ほとんどの研究では感情分析ツール VADER を使用

する.VADER は俗語、顔文字、絵文字、収縮、およびその他の自然言語属性を解釈できる。VADER は、語彙の特徴を感情の強さにマッピングする辞書を使用して、感情をポジティブまたはネガティブとして定量化する。

### 3. 実験

#### 3.1 感情と収入率の相関関係

表 1 は 7 日間の感情データと価格データを 4 つのテストで比べ極性スコアを出している。結果は収益と感情の相関関係であることを見せている。また、Granger's Causality テストで時差がある感情データが価格を予測可能かを検証する。

表 1 Performance metrics

Asset	mean	max	median	ratio
3M	0.12	0.08	0.14	0.09
Microsoft	0.18	0.08	0.12	0.13
Disney	0.17	0.08	0.16	0.11
Nike	0.05	0.07	0.06	0.0.1
Walmart	0.14	0.01	0.14	0.14

#### 3.2 モデル

本研究ではテキストから感情を予測するため Semantic Attention Model [3] を使用する。このモデルは、テキスト内の重要度の高い単語に焦点を当てることができる。また、価格予測は感情認識 LSTM モデルを使用する。LSTM の入力には、株価、いいね数、リツイート数、コメント数、ポジティブ&ネガティブ感情が使用される。そして予測した価格をもとに mean-variance モデルを使用して最適なポートフォリオを選ぶ。このモデルは収益を最大 (1)、リスクを最小 (2) を目標としている。モンテカルロ法で 50,000 個のポートフォリオ自動生成して、効率的フロンティアのポートフォリオを選択する。

$$\max \sum_{i=1}^N x_i \mu_i \quad (1)$$

$$s.t \begin{cases} \sum_{i=1}^N x_i = 1 \\ x_i \geq 0, \forall i = 1, \dots, n \end{cases} \quad (2)$$

#### 3.3 ベンチマーク

本論文では、Semantic Attention model, LSTM 価格予測, mean-variance を融合したモデルを LSTM+S と呼ぶことにする。LSTM+S の性能を評価するため従来のポートフォリオと比較する。感情認識なしの LSTM, ポートフォリオの割合を変えない Buy and Hold ポートフォリオ, t 時間の間隔で毎回一定の割合を変える Rebalancing ポート

フォリオ, すべての資産を一番収入率が良い株種目に割り当てる BestStock ポートフォリオがある。

### 4. 結果

平均的な収入率を得る Buy and Hold ポートフォリオをベンチマークにしてポートフォリオ最適化を行う。表 2 の結果で従来の手法より LSTM を使用するポートフォリオが性能が良いことがわかる。さらに LSTM+S モデルは最も良い収入率を見せている。表 3 では低下傾向の市場での精度を比較している。感情なし LSTM モデルが LSTM+S より良い収入率を得ている。この理由はネガティブラベルを十分に持っておらず、ポジティブラベルに偏りがあるためだと考えられる。

表 2 Performance metrics

Models	Capital	fAPV	BV	SR
Buy and Hold	14856.68	1.49	1.00	1.00
Best Stock	17871.20	1.79	1.20	4.17
Rebalancing	14781.47	1.48	0.99	0.91
LSTM	16966.97	1.70	1.14	3.26
LSTM + S	18047.80	1.80	1.21	1.17

表 3 Performance metrics for a down market

Models	Capital	fAPV	BV	SR
Buy and Hold	9766.90	0.98	1.00	5.27
Best Stock	9972.24	0.99	1.02	7.68
Rebalancing	9809.24	0.98	1.00	5.39
LSTM	10978.207	1.10	1.12	-14.57
LSTM + S	10702.91	1.07	1.10	-2.01

### 5. 結論

本実験はポートフォリオの最適化を目的としている。今回 VADER モデルの感情ラベルは偏りを持つことがわかった。また、感情認識ポートフォリオは従来より収入率が良く、特に低下傾向の市場に多くの収入を得た。その反面、リスク率は従来のポートフォリオより改善できなかった。実験の改善点としては、出版社によって既にラベリングされているデータを使用することが感情分析の精度面で良いと考えられる。

#### 参考文献

- [1] A. B. Paskaramoorthy, T. J. Gebbie, and T. L. van Zyl.: A framework for online investment decisions, *Investment Analysts Journal*, vol. 49, no. 3, pp. 215–231, (2020).
- [2] F. G. Ferreira, A. H. Gandomi, and R. T. Cardoso.: Artificial intelligence applied to stock market trading: A review, *IEEE Access*, vol. 9, pp. 30898–30917, (2021).
- [3] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, Image-text sentiment analysis via deep multimodal attentive fusion, *Knowledge-Based Systems*, vol. 167, pp. 26–37, (2019).