

ストラテジストレポートの感情分析と株式リターン に関する研究

The Study on the Relationship between the Sentiment Score of Strategist Reports and the Stock Returns

ニッセイアセットマネジメント株式会社 投資工学開発センター長
大学共同利用機関法人統計数理研究所 リスク解析戦略研究センター 客員教授
青山学院大学大学院 国際マネジメント研究科 客員教授
吉野 貴晶
Takaaki Yoshino

(論文要旨)

本稿は、株式ストラテジストレポートから求めた感情スコアと将来の TOPIX リターンとの関係を分析するものである。ベーシックに検討する目的からスタンダードな (1) 極性値, (2) TF-IDF, (3) Word2Vec, (4) LSTM, (5) Doc2Vec の 5 つの方法を用いている。その結果, (1) 極性値は将来のリターンとの関係が見られなかった。ストラテジストがポジティブな表現を多く使っているということだけから, 将来の株式リターンの予測は難しいことが分かる。また (2) TF-IDF も将来の株式リターンの予測が高いとは言えなかった。一方, Word の前後関係をベースとする (3) Word2Vec (W2V) と (5) Doc2Vec (D2V) や, Word の順序をモデル化の基本とする (4) LSTM は, ある程度の将来のリターンとの予測に関係していることが分かった。Word の前後関係やレポートに表れるストラテジストの特徴的な表現パターンなどが相場予測に有効となる可能性がある。

(Abstract)

This study shows the relationship between the sentiment score of strategist reports and the stock return. In this study sentiment scores are computed on the basis of five methods; (1) Polarity Value, (2) TF-IDF, (3) Word2Vec, (4) LSTM, (5) Doc2Vec. There are low correlations between these sentiment scores. And, these scores have the difference in the relationship with future stock returns. (1) Polarity Value and (2) TF-IDF do not have adequate predictability for stock returns. The reason is that number of positive/negative words do not have good relationship to future stock returns. However, (3) Word2Vec, (4) LSTM, (5) Doc2Vec provide better predictability for future stock returns. It is because the context and the pattern of expression in reports have good prediction of future stock returns.

1 はじめに

本稿は株式のストラテジストレポート¹の感情分析 (Sentiment Analysis) により求めたスコアと将来の株式リターンとの関係进行分析することを目的とする。幾つかのスタンダードな感情分析により求めた感情スコア (Sentiment Score) が将来の株式リターンとどの程度の間接関係があるのかを観察する。そして手法間で株式リターンの予測にどの程度の違いがあるかを検討して、ストラテジストレポートが株式リターンの予測に効果的であるかを考察する。

近年、自然言語処理 (NLP: Natural Language Processing) に注目が集まっている。自然言語処理とは、自然言語 (人間の言語)²を機械で処理し内容を抽出することである。自然言語の言葉が持つ意味をさまざまな方法で解析する処理の技術を指す。

自然言語処理の分野は大きく、自然言語生成 (NLG: Natural Language Generation) と自然言語理解 (NLU: Natural Language Understanding) に大別される。前者の自然言語生成とは、機械による、意味の通る自然言語の文章生成に焦点を置く技術である。入力された文書データに基づいて、文章の要約、叙述や詳説したりするものである。伝統的なものは事前のルールに基づいて文書を作成するものであった。近年は人工知能 (AI: Artificial Intelligence) の進歩により動的な文書の作成ができるようになってきている。

後者の自然言語理解は、文書に意図された意味を理解しようとする技術である。機械の読解力に焦点を置き、機械が文の意図する意味を判別できるようにするものである。

一方、テキスト (文書) を機械処理して知識発見や仮説検証を行う方法がテキストマイニングと呼ばれている。このテキストの機械処理の工程で自然言語処理が関わる。近年、投資に関して、テキストマイニングの応用が進んでいる分野が「金融テキストマイニング³」と呼ばれており、金融におけるテキストから投資に有用な情報を抽出することである。我が国においても金融テキストマイニングの研究や実務が進んでいる。

テキストマイニングは様々な目的に用いられるが、感情分析 (Sentiment Analyze) により感情スコアを求めるために行われる場合がある。感情分析とは、対象とされるテキストを調べて、そのテキストの背景にある感情としての考え方をスコア化するものである。具体的には、執筆者の考えが「ポジティブ」か「ネガティブ」、「ニュートラル」かを判断するものである。本稿ではこうした感情分析により求めた感情スコアを分析に用いている。

金融テキストマイニング研究は様々な観点から分類することができる。そのうちの 1 つは、どのような種類のテキストの情報を対象に分析するか、による分類である。「種類」に関して明確なカテゴライズは難しいが、例えば次のようにテキストの記述者の経済に関する専門性の度合いで分類することができる。

先ず (1) SNS (Social Networking Service) や掲示板などのテキスト情報を対象に分析するものがある。これらは内容が多様で経済の専門家ではない書き手も多い。しかし膨大な量のテキスト情報を集めることが可能というメリットもある。Bollen, Huina and Xiao (2011) は twitter のテキストから翌日のニューヨークダウ平均株価を予測する研究を行った。気分解析

ツールである Google-Profile of Mood States (GPOMS) を利用して求めた感情スコアとダウ平均株価が相関関係にあることを示している。そしてダウ平均株価を 86.7%の精度で予測できるとしている。

次に (2) オンラインの経済ニュースなど専門的なテキスト情報を対象とする分析である。日経 QUICK ニュース⁴、ロイターニュース⁵や Bloomberg ニュース⁶などは経済の専門的な知識がある記者のテキスト情報で、機関投資家などが取引の参考に使っているものである。五島・高橋 (2017) は日経 QUICK ニュースを使って極性辞書の作成を行っている。極性辞書とは簡単には、ある単語がネガティブなのか、ポジティブなのかを、-1 (ネガティブ) から 1 (ポジティブ) までのスコアの形で表現するものである。五島・高橋 (2017) ではニュース記事配信日の株式リターンを用いて極性辞書を作成している。そして同辞書を用いて将来のニュース記事の感情スコアを求めると、同スコアがニュース記事前後の株式リターンと正の関係があることを示しており、経済・金融分野に特化していない一般的な極性辞書よりも効果的に分類できていることを示している。

最後が (3) 専門的なテキスト情報として、経済の専門家が記述する経済レポートを対象とする分析である。和泉・後藤・松井 (2011) では、日本銀行が発行する金融経済月報が実際の市場動向をどの程度説明しているのかについて検証を行った。金融経済月報のテキスト情報から月末の債券価格を予想する運用テストを行った結果、既存の計量経済モデルと比べて、安定して、ほぼ最高水準の運用益をあげることができることを示している。

本稿は証券会社の株式ストラテジストのレポートを対象とする分析である。経済の専門性の度合いによる分類では、最後に取り上げた「(3)」に該当する。株式ストラテジストは証券会社の調査部門に所属するが、同じ調査部門には企業アナリストも属している。このアナリストレポートを対象とする分析は我が国でも広く行われている。工藤・永島・宮崎 (2017) は、アナリストレポートから個別株式のセンチメントスコアを求めて、それらの違いがその後の超過収益率に影響していることを明らかにしている。

このような企業アナリストレポート解析の背後には次の根拠がある。投資家はアナリストの将来の利益予想に注目しているため、業績の予想値が修正されると株価も反応する。しかし近年、情報の効率性が高まりアナリストが業績予想を修正すると株価が瞬時に反応するため、修正後の投資では超過リターンの確保が難しい。そこでアナリストが業績予想を修正する前に、レポートの文脈から将来の業績予想の姿勢を捉えようとするのである。数値情報である業績予想の修正は、投資家へのインパクトも大きいので、アナリストは確信を高めてから踏み切る傾向がある。企業レポートの自然言語処理による得点化は、将来のアナリストの利益予想修正の予兆を捉える目的がある。

一方で本稿が分析対象とする一般にストラテジストレポートを使った自然言語処理の評価でのモデル化は見られない。これは、①ストラテジストが相場水準の見通しの修正しても株式市場に大きな影響を及ぼすとは考え難いこと、②レポートの文脈で将来の相場水準変更の予兆を捉えても有用性に乏しいと見られること、そして③将来の相場予想水準の変更

が事前の文脈に表れているとは考え難いことなどがある。また④ストラテジストは各々が独特の表現を持つため、学習サンプルの扱いが難しいこともある。

しかしストラテジストは市場の専門家で、知見も豊富に持っており、市場全体のセンチメントを敏感に捉えている可能性がある。こうした傾向がレポートの文脈に表れることで市場予想が可能となるかもしれない。そこで本稿はストラテジストレポートの解析により相場予測スコアを求めて、株式市場の予測可能性を検証する。

本稿で取り上げた感情分析の方法はスタンダードなものとする。そして、それぞれの方法の特徴と比較を検討する。近年は様々な感情分析の方法が開発されているが、シンプルな方法を使うことで、株式ストラテジストレポートにどの程度の予測力があるのかをベーシックに捉えるためである。

本稿の構成は次の通りである。まず第2節では分析データを示す、ここではリターンとの関係を捉える検証方法も確認する。第3節では本稿で分析対象とする感情分析を示す。ここでは本稿で感情分析に用いる手法を解説する。第4節で分析結果を示す。第5節のおわりにでは今後の課題などを示す。

2 データと検証方法

本稿は株式ストラテジストレポートを用いて幾つかの感情分析の方法により計算した「感情スコア」が、将来の株式リターンとどの程度の関係があるかを分析するものである。

感情スコアの算出方法に関しては、3節で示す。本節では感情スコアを計算する際に分析に用いるデータと、株式リターンとの関係の検証データを示す。

本稿で使用する株式ストラテジストレポートは、ある特定の証券会社の1社のみとする。ストラテジストには特有の文章の癖があるため、個別毎に評価することが妥当と考えるからである。また、分析にはレポートの1ページのみ利用する。1ページ目には最も重要な主張やコメントが要約されて記載されているからである。レポートによっては数ページから数十ページのものボリュームが様々である。また株式相場の予測のコメント以外に、例えば環境問題など様々なトピックに関する解説などが含まれているものがある。このため、レポートの内容全てを扱うのでは株式相場の予測を目的とする感情スコアの算出にそぐわない面もあると考えられるからである。

学習データに関しては2012年9月～2019年9月の期間に発行された2206本のレポートを対象とする。学習データでモデル化を行った後、そのモデルを用いたテスト期間として2019年10月～2020年9月までの1年間のレポート(342本)の感情スコアを算出して、それぞれのレポートのスコアがその後の株式相場全体の変動をどの程度予見できるかを検証する。モデル化する上で学習サンプル数が十分な量とは言えないが、実務的に利用可能なサンプルでどの程度の相場予測が可能かも検討する。そしてレポートから求める感情スコアとの関係を捉える株式相場に関しては配当込みTOPIX(東証株価指数)を用いる。

3 分析対象とする感情分析の方法について

本稿で、とりあげる感情分析はそれぞれ次の5つの方法をベースに求めている。これらの方法を用いて求めた感情スコアと将来の配当込み TOPIX の変動との関係を分析する。

- (1) 極性値
- (2) TF-IDF
- (3) Word2Vec (W2V)
- (4) LSTM (Long Short-Term Memory)
- (5) Doc2Vec (D2V)

以下では、それぞれの手法に関する概要と実際のデータの使い方を示す。

3.1 極性値を用いた感情スコアの算出

予測スコアの作成手法として分かり易い方法が、極性辞書を用いるものである。

レポートに使われているワードに関して、ポジティブとネガティブワードを、その程度である「極性値」を使って集計する方法である。極性値は学習レポートを使って分析者が作ることも可能である。しかし本稿では、東京大学 和泉・坂地研究室で開発、公開されて誰でも取得可能な金融極性辞書⁷を用いる。分析対象の学習サンプルが十分でないことに対応するためである。近年はAIのコモディティ化⁸の流れにあるなかで、辞書に関しては分析者が機械学習をしなくてもモデル化が可能となる良く知られた例に当てはまる。例えば、同辞書で「増大」に関する極性値は0.49である、対義語となる「減少」は-0.71となっている。ポジティブがプラスとなり、ネガティブがマイナスである。

この辞書を用いてテスト期間の2019年10月～2020年9月までの1年間のレポートを用いて、次の分析を行う。今回の分析では既に存在する金融極性辞書を用いているため、テスト期間に存在するレポートのみを評価する。

まず、それぞれのレポートの分かち書きを行う。分かち書きにはMeCabを用いている。そして名詞、動詞と形容詞のみを抽出する。

そして、各レポート内に出現するキーワードを計数し、極性を表す重みで掛け合わせた値を、計数したキーワード数で割ることでレポートの感情スコアを算出する。これは五島・高橋(2017)で用いられているものである(1)式により、それぞれの極性値を各レポートにおける出現数で加重合計したものを、それぞれのレポートの感情スコアとする。

$$\text{ストラテジストレポートの感情スコア} = \frac{\sum_{k=1}^n F_k \cdot \text{Weight}_k}{\sum_{k=1}^n F_k} \quad (1)$$

ここで F_k は金融極性辞書の k 番目のキーワードが対象となるストラテジストレポート内に出現した頻度⁹, $Weight_k$ は k 番目のキーワードの極性度合いを表す数値, n は金融極性辞書に定義されているキーワード数を表す。

3.2 TF-IDF を用いた感情スコアの算出

テキストにおける重要語の抽出を行うツールで広く使われる手法が TF-IDF である。Term Frequency (TF) と、Inverse Document Frequency (IDF) の2つの尺度を掛け合わせてつかうことから、こう呼ばれている。

TF は情報検索などでよく用いられるもので、テキスト内で「出現回数の多い単語」を拾い、「出現回数の多い単語」ほど「重要な単語」であると評価するものである。具体的には次のように求められる。

$$TF = \frac{\text{文書 A における単語 X の出現回数}}{\text{文書 A におけるすべての単語の出現回数の和}} \quad (2)$$

IDF は、ある単語が出てくる文書頻度の逆数である。多くの文書中に登場する単語は、一つの文書の特徴語にはなりにくいと捉える。

$$IDF = \log \left(\frac{\text{総文書数}}{\text{単語 X が出現する文書数}} \right) \quad (3)$$

そして、TF-IDF の値を (4) 式で定義される。

$$TF-IDF = TF \times IDF \quad (4)$$

単語 X の出現頻度 TF に、多くの文書に出現すればするほど値が小さくなる逆文書頻度の IDF を掛けることで求められる。これにより、文書の中で出現回数が多く、かつ他の文書の中であまり出現しない単語が抽出される。

ここで (3) 式 of IDF の対数の底としては 2 が用いられるのが一般的である。また対数が取られる理由には、TF-IDF の計算において IDF における総文書数の値の変化における寄与を小さくするためである。

TF-IDF を使った感情スコアの計算は次のようである。

先ず、学習データの 2012 年 9 月～2019 年 9 月 (2206 本) とテストデータの 2019 年 10 月～2020 年 9 月 (342 本) に発行された合計 2548 本のレポートを対象に分ち書きを行う。さらに各レポートにおいて助詞などを除き、名詞、動詞と形容詞のみの Word を抽出する。ここまでは前述の「極性値」を用いた手法と同じである。

次に、それぞれのレポートにおける Word の TF-IDF を計算する。(2)式における文書 A には、それぞれ発行されるレポートが1つずつ入る。単語 X は 2548 本のレポートにおいて、先に抽出した Word である。合計 13226 個の word が該当する。従って、それぞれのレポートについて、13226 個の word のそれぞれに付された TF-IDF が求められる。データとしては、それぞれのレポートに 1 行 13226 列のベクトルとして TF-IDF のデータセット (以後、TF-IDF ベクトルと表す) が計算される。

次に学習期間における 2206 本の TF-IDF ベクトルを、レポート発行後 TOPIX が上昇したサンプルのベクトルと、下落したサンプルのベクトルの 2 つに分類する。具体的にはレポート発行日の翌日から 5 日間の TOPIX の上昇と下落で分ける。この結果、上昇した TF-IDF ベクトルは 1316 本、下落した TF-IDF ベクトルは 890 本となる。

次にテスト期間における 342 本の TF-IDF ベクトルから、その後の相場の上昇、下落予測のベースとなる感情スコアの計算を行う。実際には、テスト期間におけるそれぞれのレポートに対応した TF-IDF ベクトルと、学習期間の 2206 本の TF-IDF ベクトルの間で総当たりでコサイン類似度を計算する。そして、各テスト期間のレポートについて、学習期間で上昇サンプルの 1316 本のコサイン類似度の平均値から、下落サンプルの 890 本のコサイン類似度の平均値を引いた値を、テスト期間のレポートの感情スコアとする。

3.3 Word2Vec (W2V) を用いた感情スコアの算出

テキスト中の Word について実数値を要素に持つベクトルで表したものは「Word の分散表現」と呼ばれている。文章中の単語の前後関係からニューラルネットワークを用いて Word を分散表現に変換する手法が Word2Vec である。Mikolov, Chen, Corrado and Dean (2013) により考案されたもので、広く一般的に Word の分散表現への変換に用いられるものである。

実際に Word2Vec を使った感情スコアの計算は次のようである。

まず、学習データの 2012 年 9 月～2019 年 9 月 (2206 本) とテストデータの 2019 年 10 月～2020 年 9 月 (342 本) に発行された合計 2548 本のレポートを対象に分かち書きを行う。さらに各レポートにおいて助詞などを除き、名詞、動詞と形容詞のみの Word を抽出する。ここまでは前述までの 2 つの方法と同じである。

次に、合計 2548 本の全てのレポートを使って Word の分散表現を計算する。実際には python のライブラリである gensim を用いる。今回は 200 次元、CBOW で取得する。また、Word の前後関係を見る際の前後はそれぞれ 3 Word と設定している。そしてそれぞれの Word の分散表現のベクトルは長さが 1 となるように基準化している。

次に、それぞれのレポートに使われている Word に対応する分散表現のベクトルを、レポート毎に和をとる。これがレポート全体で見た分散表現とする..

次に学習期間における 2206 本の分散表現のベクトルを、レポート発行後 TOPIX が上昇したサンプルの分散表現のベクトルと、下落したサンプルの分散表現のベクトルの 2 つに分類する。具体的にはレポート発行日の翌日から 5 日間の TOPIX の上昇と下落で分ける。

この結果、上昇した分散表現のベクトルは 1316 本、下落した分散表現のベクトルは 890 本となる。

次にテスト期間（2019 年 10 月～2020 年 9 月）における 342 本の分散表現のベクトルから、その後の相場の上昇、下落予測のベースとなる感情スコアの計算を行う。実際には、テスト期間におけるそれぞれのレポートに対応した分散表現のベクトルと、学習期間の 2206 本の分散表現のベクトルの間で総当たりでコサイン類似度を計算する。そして、各テスト期間のレポートについて、学習期間で上昇サンプルの 1316 本のコサイン類似度の平均値から、下落サンプルの 890 本のコサイン類似度の平均値を引いた値を、テスト期間のレポートの感情スコアとする。

3.4 LSTM を用いた感情スコアの算出

Word2Vec は Word の前後との関係を使って、対象の Word の分散表現を行う。しかし実際に人間が文書を読む際には、前の Word の理解に基づいて、その後の単語の理解を行う。RNN（Recurrent Neural Network）はこうした点に対処するもので、前の情報を持続されるものである。基本的な構造は、現在入力値と前回の出力値を合計し、活性化関数として tanh（Hyperbolic Tangent Function：双曲線正接関数）を適用して出力するものである。しかし、何度も重みと活性化関数として tanh の微分を掛け合わせることから、勾配消失、或いは爆発問題が発生する。これが長期依存性の問題として指摘される。こうした実態にも対処する手法が LSTM（Long Short-Term Memory）である。この方法は過去のデータを tanh ではなく線形和で保持するため、勾配が極端に大きくなったり小さくなったりする問題に対応する。

本稿では LSTM を用いたモデルがどの程度、将来のリターンの予測力があるのかを分析する。具体的には python のライブラリである PyTorch を利用して次のように行う。

まず、学習データの 2012 年 9 月～2019 年 9 月（2206 本）とテストデータの 2019 年 10 月～2020 年 9 月（342 本）に発行された合計 2548 本のレポートを対象に分かち書きを行う。さらに各レポートにおいて助詞などを除き、名詞、動詞と形容詞のみの Word を抽出する。ここまでは、これまで紹介した方法と同様である。

次に、学習データの 2206 本のレポートのそれぞれについて、一連の Word のセットを、個々の Word に対応する分散表現のベクトルに変換する。これは Word2Vec で求めたものを用いる¹⁰。分散表現のベクトルは 200 個の要素のものを用いている。これをインプットデータとする。一方、レポート発行後から配当込み TOPIX が上昇したか、下落したかの 2 つのカテゴリの何れかに該当するかを分類先のカテゴリとする。レポート発行日の翌日から 5 日間の配当込み TOPIX の上昇と下落で分ける。この結果、上昇した分散表現のベクトルは 1316 本、下落した分散表現のベクトルは 890 本となる。モデル構築に関しては、損失関数のインプットには、対数 softmax 関数を用いる。またパラメータ推計に関してエポックは 30 回とする。

次にテスト期間に関するデータは次のように扱う。構築されたモデルにテスト期間にお

けるそれぞれのレポートに対応した分散表現のベクトル系列をインプットする。そして、各テスト期間のレポートについて、上昇する確率から下落する確率を減じたものを、テスト期間のレポートの感情スコアとする。

3.5 Doc2Vec (D2V) を用いた感情スコアの算出

Doc2Vec は文書を1つの分散表現のベクトルに変換する手法の1である。Le and Mikolov (2014) で考案されたものである。Doc2Vec は Word2Vec のアルゴリズムを、文、段落、文章などの連続する表現に拡張したものである。前述の Word2Vec も文書を1つの分散表現のベクトルに変換するものであるが、こちらは第1段階で Word の分散表現を求めた後に、第2段階として文書内に出現する Word の分散表現を合計する2段階である。

Doc2Vec も Word2Vec と同様に文書内の Word の前後関係がベースであるが、このような2ステップではない。本稿では PV-DM (Distributed Memory Model of Paragraph Vector) を用いて、文書を分散表現に変換する方法を用いる。

Doc2Vec を使った感情スコアの計算は次のようである。

まず、学習データの2012年9月～2019年9月(2206本)とテストデータの2019年10月～2020年9月(342本)に発行された合計2548本のレポートを対象に分ち書きを行う。さらに各レポートにおいて助詞などを除き、名詞、動詞と形容詞のみの Word を抽出する。ここまでは前述の方法と同様である。

次に、合計2548本の全てのレポートを使って文書の分散表現を計算する。実際には python のライブラリである gensim を用いる。今回は300次元で取得する。また、Word の前後関係を見る際の前後はそれぞれ8 Word と設定している (gensim のデフォルトを利用)。

次に学習期間における2206本の分散表現のベクトルを、レポート発行後 TOPIX が上昇したサンプルの分散表現のベクトルと、下落したサンプルの分散表現のベクトルの2つに分類する。具体的にはレポート発行日の翌日から5日間の配当込み TOPIX の上昇と下落で分ける。この結果、上昇した分散表現のベクトルは1316本、下落した分散表現のベクトルは890本となる。

次に上昇と下落のそれぞれのカテゴリーで文書の分散表現のベクトルを平均する。上昇した分散表現のベクトルは1316本、下落した分散表現のベクトルは890本のそれぞれを平均している。そして、それぞれを「ポジティブ基準ベクトル」「ネガティブ基準ベクトル」と呼ぶ。

次にテスト期間における342本の分散表現のベクトルから、その後の相場の上昇、下落予測のベースとなる感情スコアの計算を行う。実際には、テスト期間におけるそれぞれのレポートに対応した分散表現の数値ベクトルと、2つの基準ベクトルであるポジティブ基準ベクトル、ネガティブ基準ベクトルとの間でコサイン類似度を計算する。こうした処理で各レポート毎に、ポジティブ、ネガティブ基準ベクトルに対する類似の程度が求められる。

最後に、ポジティブ基準ベクトルからの類似の程度から、ネガティブ基準ベクトルからの

類似の程度を減じたものが感情スコアとする。

3.6 感情スコア間の違いについて

本節では (1) 極性値, (2) TF-IDF, (3) Word2Vec (W2V), (4) LSTM, (5) Doc2Vec (D2V) の 5 つの方法で求めた感情スコアがどのように違いがあるかに関して相関係数で概観する。テストデータの 2019 年 10 月～2020 年 9 月 (342 本) に発行されたレポートを対象にそれぞれの手法で求めた感情スコアの相関係数を示す (ピアソンの積率相関係数)。

表 1 感情スコアの相関係数行列

	(1)極性値	(2)TF-IDF	(3)W2V	(4)LSTM	(5)D2V
(1)極性値	1.00				
(2)TF-IDF	0.14	1.00			
(3)W2V	0.24	0.58	1.00		
(4)LSTM	0.01	0.02	-0.04	1.00	
(5)D2V	0.08	0.33	0.19	0.05	1.00

(注) 2019 年 10 月～2020 年 9 月 (342 本) に発行されたレポートを対象に 5 つの手法それぞれで求めた感情スコア間のピアソン積率相関係数。

(出所) 該当のストラテジストレポートをもとに筆者作成

全体的に 5 つの手法による感情分析の値の間で求めた相関係数の値が高くない。特に、(3) Word2Vec と (4) LSTM との間で相関係数が -0.04 とわずかながら負となった。手法間で傾向が異なるということは、感情分析を相場予測に利用する上で手法の選び方が重要であることを示す。

相関係数が最も高かったのは (2) TF-IDF と (3) Word2Vec との間である (相関係数:0.58)。TF-IDF の手法は Bag of Words が基本となっている。すなわち単語の出現頻度を数えるものである。TF-IDF は、他の文書には表れ難いが、対象となる文書に頻繁に現れる Word の得点を高める形で改良された手法である。これに対して (3) Word2Vec を用いた感情分析では、Word の分散表現のベクトルの和を計算するので、対象となる文書において出現頻度が多い Word のベクトルの和の部分が大きくなる。こうした傾向が表れていると考えられる。

とりわけ (4) LSTM は他の手法との相関係数が低い。最も相関係数が高い関係でも (5) Doc2Vec との 0.05 に過ぎない。LSTM は文書における単語のつながりを学習しているため、他の手法とは異なる点が多いのだろう。

4 分析結果

分析結果に関しては、(1) 極性値、(2) TF-IDF、(3) Word2Vec (W2V)、(4) LSTM、(5) Doc2Vec (D2V) の5つの方法で算出した感情スコアと将来の配当込み TOPIX リターンとの関係を観察する。

結果は表2に示している。テスト期間となる2019年10月～2020年9月に発行された342本のレポートに関して、5つの感情スコアに関して、事前に「上昇」、「中立」、「下落」と3つの予測の何れかに該当するかを求める。ここでは5つの感情スコアに関して、3つのカテゴリに概ね同数のサンプルが入るようにそれぞれの閾値を決める。

そして、レポートの発行後の翌営業日から5営業日間の配当込み TOPIX の収益率との関係を捉える。3つのカテゴリのそれぞれで配当込み TOPIX の平均値を求める。また、それぞれのカテゴリのなかのサンプルにおいて相場が上昇した割合を勝率とする。

更に10営業日後までの配当込み TOPIX の収益率でも同様の計算を行う。

表2 相場予測スコアの検証結果

事前 予測	5日後 平均値	勝率	10日後 平均値	勝率	サンプル 数
(1)極性値					
上昇	0.27%	57.4%	0.87%	63.5%	115
中立	<u>-0.05%</u>	<u>45.2%</u>	<u>0.06%</u>	<u>54.8%</u>	115
下落	0.44%	62.5%	0.09%	61.6%	112
(2)TF-IDF					
上昇	0.18%	55.4%	0.38%	<u>57.4%</u>	101
中立	<u>0.06%</u>	<u>49.2%</u>	<u>0.26%</u>	58.7%	126
下落	0.42%	60.9%	0.41%	64.3%	115
(3)W2V					
上昇	0.54%	55.4%	0.66%	57.9%	121
中立	-0.16%	55.0%	0.35%	62.2%	111
下落	0.24%	54.5%	-0.01%	60.9%	110
(4)LSTM					
上昇	0.14%	61.1%	<u>0.29%</u>	<u>55.6%</u>	108
中立	0.42%	55.6%	0.40%	65.3%	124
下落	<u>0.06%</u>	<u>48.2%</u>	0.33%	58.2%	110
(5)D2V					
上昇	0.37%	55.6%	0.47%	63.9%	108
中立	<u>-0.01%</u>	55.3%	0.34%	<u>58.3%</u>	132
下落	0.35%	<u>53.9%</u>	<u>0.22%</u>	58.8%	102
全サンプル	0.22%	55.0%	0.35%	59.9%	342

(注1) 2019年10月～2020年9月までの1年間のレポートを対象とする。勝率はそれぞれの予測に対して相場が上昇した割合。検証に用いるリターンは配当込み TOPIX を用いる。

(注2) 表中太字は上昇、中立、下落の3つのカテゴリで比較して最大。斜体下線は最小。

(注3) 事前に上昇と予測したケースと事前に下落と予測したケースのそれぞれの5日後のリターンの平均値の差の検定をした結果、(1)から(5)の全てのケースにおいて両側有意水準10%で有意に離れていない。また、10日後のケースでも同様に有意に離れていない。

(出所) 東京証券取引所のデータをもとに筆者作成

ストラテジストレポートの感情分析と株式リターン に関する研究

(1) 極性値を使った相場予測スコアは相場予測精度が高くない。5日後に「上昇」すると事前に予測していたサンプルが 115 レポートであったが、実際にそれらのレポートの翌日から5日間の配当込み TOPIX のリターンを平均すると 0.27%であった。しかし事前に「下落」と予測している場合には 0.44%となっており、事前に下落予測をしていた方が上回っている。また勝率も事前の下落予測の方が上昇予測よりも高い。

10日後の結果も同様に示しているが、事前に「上昇」と予測していた場合の平均リターンが 0.87%と高かったものの、「下落」と予測していた場合の平均リターンも 0.09%となり「中立」の-0.06を上回っており、妥当な予測ができているとは言えないだろう。ストラテジストレポートを使った予測に関しては、ポジティブ、或いはネガティブの Word の極性値を単純に集計するだけでは将来の相場予測が難しい傾向があることがわかる。

Bag of Words をベースとする (2) TF-IDF の手法も事前の予測精度が高くはない。モデルの構築の際に用いる学習サンプルでは5日後までの配当込み TOPIX リターンを用いている。このため、同じ5日後までのリターンの予測精度にとりわけ着目すべきだろう。しかし事前の「下落」予測のサンプルに関して5日後までの配当込み TOPIX リターンの平均値は 0.42%、勝率は 60.9%となり「上昇」「中立」と比べて高かった。

一方、(3) Word2Vec (W2V) の手法は、前述した2つの手法と比べるとある程度の予測精度の高さが見られる。5日後の結果を見ると、事前に「上昇」と予測していた場合の平均リターンが 0.54%と高く、「中立」や「下落」と予測していた場合の平均リターンを上回っている。勝率に関しても、同様に 55.4%と「中立」や「下落」と比べて高い。10日後リターンに関しても平均値は「上昇」が最大だった。Word2Vec は文書中の Word の前後関係から分散表現のベクトルに変換された値をベースとしている。レポートに表れるストラテジストの特徴的なクセなどの特徴が把握できることが相場予測に有効となるかもしれない。

(4) LSTM も5日後までのリターンの予測の観点では、一定の評価ができる手法である。事前に「上昇」と予測していた場合の勝率は「中立」や「下落」と比べて高く 61.1%と最大である。また、最小の勝率が「下落」の 48.2%となっている。LSTM では Word の順序をモデル化の基本としていることや、今回は LSTM へのインプットに、Word 自体でなく Word2Vec による分散表現ベクトルを用いていることから、(3) Word2Vec の手法を更に改良する形となっている。

Word の前後関係から文書全体の分散表現のベクトルを推計する (5) Doc2Vec (D2V) も予測力の観点で一定の効果が確認される。前に「上昇」と予測していた場合の平均値と勝率は共に、「中立」や「下落」と比べて高く 0.37%、55.6%であった。また、10日後までの予測でも、「上昇」と予測していた場合の平均値と勝率は共に最大であった。

しかし、これらの分析結果に関しては注意点もある。事前に上昇と予測したケースと事前に下落と予測したケースのそれぞれの5日後のリターンの平均値の差の検定をした結果、(1)から(5)の全ての手法のケースにおいて両側有意水準 10%で有意に離れていない。また、10日後のケースでも同様に有意でない結果が示されていることは留意すべきである。予測

結果に一定の効果がみられるものもあるが、統計的な観点では有意とまでは至らない。

5 おわりに

本稿は、株式ストラテジストレポートを使って (1) 極性値, (2) TF-IDF, (3) Word2Vec (W2V), (4) LSTM, (5) Doc2Vec (D2V) の5つの方法で算出した感情スコアと将来の配当込み TOPIX リターンとの関係を分析した。

本稿で対象としたレポートを使った分析結果では、(1) 極性値は、あまり将来のリターンとの関係が見られなかった。ストラテジストがシンプルにレポートにポジティブな表現を多く使っているということだけから、将来の株式リターンの予測は難しいことが分かる。

Bag of Words をベースとする手法の (2) TF-IDF も将来の株式リターンの予測が高いとは言えなかった。過去 (学習期間) で発行後の株式リターンが上昇 (下落) することで分類しても、それが将来も同様の傾向とはならない。

一方で、Word の前後関係をベースとする (3) Word2Vec (W2V) と (5) Doc2Vec (D2V) や、Word の順序をモデル化の基本とする (4) LSTM は、ある程度の将来のリターンとの予測に関係していることが分かった。Word の前後関係やレポートに表れるストラテジストの特徴的な表現パターンが相場予測に有効となるかもしれない。

今回の分析ではスタンダードな手法を使った感情スコアを算出した。これはベーシックな手法で株式ストラテジストレポートが将来のリターン予測に有効であるのかを検討する目的があるからである。手法間で感情スコアが大きく異なることから、手法の選択が重要であるなかで、統計的な有効性は確認できなかったものの、一定の予測力の可能性を示すものが見られたことは注目できる。実務面で応用する上では、過学習を避けつつモデル化におけるハイパーパラメータの調整が必要となるだろう。

近年は Devlin, Chang, Lee and Toutanova (2019) の BERT に代表される自然言語処理の手法の開発も進んでいる。今後はこうした手法も含めたモデル精度の比較検討を行いたい。また、今回はある1社のレポートのみを分析対象としたケースを紹介した。複数社のレポートのそれぞれの分析を行い、それらを融合する活用も相場予測に効果的となる可能性があるため、今後の課題としたい。

本稿を作成するにあたり、青山学院大学大学院国際マネジメント研究科 森田充教授、ニッセイアセット投資工学開発部のメンバーより有益なコメントを頂いた。ここに記し、感謝申し上げます。本稿の内容は筆者らが所属する組織を代表するものではなく、全て個人的な見解である。また、当然のことながら、本稿における誤りは全て筆者の責に帰するものである。

(参考文献)

- 和泉潔・後藤卓・松井藤五郎 (2011) 「経済テキスト情報を用いた長期的な市場動向推定」『情報処理学会論文誌』52 (12) ,3309-3315
- 工藤秀明・永島淳・宮崎義弘 (2017) 「自然言語処理技術を用いたアナリスト・レポートの実証分析」『証券アナリストジャーナル』, 55 (9) ,66-77
- 五島圭一・高橋大志 (2017) 「株式価格情報を用いた金融極性辞書の作成」『自然言語処理』, 24 (4) ,547-577
- 関和広・水田孝信・八木勲・落合友四郎・酒井浩之・和泉潔 (2017) 「金融情報学：ファイナンスにおける人工知能応用」『人工知能』32 (6) , 905-910
- Bollen, J., M. Huina, and X. Zeng (2011) “Twitter Mood Predicts the Stock Market”, *Journal of Computational Science*, 2 (1) ,1-8.
- Devlin, J., M. Chang, K. Lee, and K. Toutanova (2019) “BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding ” , *arXiv preprint arXiv:1810.04805*.
- Ito, T., H. Sakaji, K. Tsubouchi, K. Izumi, and T. Yamashita (2018) “Text – Visualizing Neural Network Model : Understanding Online Financial Textual Data” In : Phung, D., V. Tseng, G. Webb, B. Ho, M. Ganji, and L. Rashidi (eds), *Advances in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science*, Springer, 10939, 247–259.
- Le, Q., and T. Mikolov (2014) ”Distributed Representations of Sentences and Documents”, *CoRR*, abs/1405. 4053, 1-9.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013) “Efficient Estimation of Word Representations in Vector Space ”, *arXiv preprint arXiv: 1301.3781*, 1-12.

1 株式ストラテジストは、株式の投資戦略を設計する立案者のことである。一般に経済・産業・企業の動向、需給要因まで、様々な視点で投資環境を分析して、投資方針を提供する。

2 「自然言語」は、コンピュータで用いる「プログラミング言語」から、人間が日常用いている言語を区別する目的で用いられる。自然言語とは、通常我々人間が用いている日本語や英語などのこと。

3 「金融テキストマイニング」という表記は一般的でもないが、関・水田・八木・落合・酒井・和泉 (2017) で同表記が使われている。人工知能学会で認知された表現となる。

4 日経 QUICK ニュースは、日本経済新聞社と QUICK 社によって投資家向けに専用の端末を通じて配信されるニュース。

5 トムソンロイター社が配信しているニュース。

6 ブルームバーグ (Bloomberg L.P.) が配信しているニュース。

7 <https://sites.google.com/socsim.org/izumi-lab/tools/>金融専門極性辞書

8 AI のコモディティ化とは、AI 自体は誰もがアクセスできる「ライブラリ」という形で提供されることから、既に学習済みのデータが利用できることなど、幅広い意味で用いら

れる。

9 対象となるレポート中に出現しなかった **Word** は該当する F が 0 回となる。

10 分散表現に関しては, **pytouch** で提供されている学習済みのベクトルを用いることも可能である。しかしストラテジストレポート特有の **Word** の分散表現を用いることが, より妥当と考えた。