

アナリ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 Analyst 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、企業アナリストのレポートの感情スコアと将来の個別銘柄の株式リターンとの関係を分析することを目的とする。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, (4) LSTM を除いた 4 手法は、感情ファクターが銘柄選択にある程度、有効に機能する傾向が見られた。特に (1) 極性値は銘柄選択効果が最も高かった。今回はレポートの要約部分となる表紙ページのみを分析対象としたため、複雑なモデルと比べてシンプルな (1) 極性値の集計が、将来のリターンとの関係でより妥当な感情スコアが求められたと考えられる。一方, (5) Doc2Vec (D2V) も銘柄選択効果が高かった。レポートの表紙部分の Word の前後関係に関するアナリストの表現のクセなどの特徴を捉えられると、銘柄選択に有効となる可能性がある。手法間で感情スコアの傾向が大きく異なることや、将来のリターンとの有効性も異なることから、実務に利用する際には手法の選択が重要であることが分かった。

(Abstract)

This study shows the relationship between the sentiment score of Analyst reports and the individual future stock returns. 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, Except for LSTM, these scores have the relationship with future stock returns. Especially (1) Polarity Value has significant predictability for stock returns. The number of positive/negative words in summary page has good relationship to future stock returns. On the other hand, (5) Doc2Vec provide adequate 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、セルサイド¹⁾に所属する企業アナリストのレポート²⁾（以後、アナリストレポートと表す。）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を分析することを目的とする。リターンとの関係の検証に関しては、運用実務への示唆を目的として、3分位ポートフォリオの手法を用いる。そして、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あるのかを観察して、感情分析が個別株式のリターン予測に効果的であるかを考察する。吉野（2021）では、ストラテジストレポート³⁾から求めた感情スコアと株式市場全体の変動を示すTOPIX（東証株価指数）リターンとの関係を明らかにした。分析の結果、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、感情スコアの算出手法によっては、一定の予測力の可能性を示すものが見られた。本稿では同様の感情分析の手法を用いて、アナリストのレポートと個別株式リターンとの関係を観察する。

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）⁴⁾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。自然言語処理は様々な分野で活用されている。身近なものが「検索エンジン」である。利用者が入力したワード（Word）が使われている記事を検索するという基本的なものから、近年は発展が見られている。例えば、利用者が入力したワードに関連する記事の検索など、利用者の意図に基づいて、関連性の高い結果が表示できるようになっている。自然言語処理の他の実用例としては、別の言語から利用者の母国語に翻訳する「言語翻訳」や、音声パターンを認識して意味を推測し、有用な回答を提供する「スマートアシスタント⁵⁾」などもある。

こうした自然言語処理では、統計や機械学習の手法を使用して、非構造化テキストのデータを意味のある分析用のデータに変換するテキスト（文書）分析が行われる。具体的にテキスト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。近年、投資の世界でも、テキストマイニングの応用が進んでおり「金融テキストマイニング⁶⁾」と呼ばれている。金融の分野で利用されるテキストから投資に有用な情報を抽出するものである。代表的なものに、分析対象とされる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「感情分析」がある。レポートや記事などの執筆者の考えを「ポジティブ」か「ネガティブ」、「ニュートラル」かのスコアに変換するものである。本稿ではアナリストレポートの感情分析から求めた感情スコアと株式リターンとの関係を分析する。

金融の分野で利用されるテキストを使った感情分析と株式リターンに関する全般的な先行研究のレビューは吉野（2021）で示されている。本稿では、研究の範囲として設定するアナリストレポートの感情スコアと、個別株式リターンとの関係に関する先行研究をレビューする。

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

実際には、金融関連のニュースや SNS の感情分析を行ったスコアとリターンとの関係に関する研究に比べると、アナリストレポートの感情分析を行ったスコアに関する先行研究は限られている。これはニュースなどと比べて広範囲にアナリストレポートが取得できる研究者⁷が限定されているからかもしれない。

米国株を対象とした研究は次のようである。Twedt and Rees (2012) はアナリストレポートの TONE (語調) を求めて、当該レポートで示されるアナリスト予想 EPS の市場コンセンサスとの乖離など、レポートの掲載される定量データをコントロールしても、レポート発行前後の合計 3 日間のリターンに対して、TONE が有意に影響を与えていることを示した。Twedt and Rees (2012) で用いられた TONE は (ポジティブワード数－ネガティブワード数) ÷ 全ワード数という Bag-of-Words (単語の出現回数) をベースとしたシンプルなものであった。Huang et al. (2014) はアナリストレポートを構成する文 (Sentence) をポジティブとニュートラル、ネガティブに分けて、(ポジティブセンテンス数－ネガティブセンテンス数) ÷ 全センテンス数の指標を OPN と定義した。そして、レポートに示されているアナリストの推奨レーティングなどをコントロールしても、OPN がレポート発行日と翌日の 2 日間のリターンと有意な関係があることを示した。Miwa (2021) は Twedt and Rees (2012) で用いられていた TONE の変化である Δ TONE が、レポート発行当日と翌日の 2 日間だけでなく、10 日後まで、リターンに対して有意に影響を与えることを示した。

一方、日本市場における研究では、工藤・永島・宮崎 (2017) は、アナリストレポートから個別株式の感情スコアを求めて、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。そして平松ほか (2021) では、Twedt and Rees (2012) の TONE を日本株のアナリストレポートを使って算出して、楽観的なトーンのレポートの方が、悲観的なトーンのレポートよりレポート発行時のリターンに与える影響が大きくなることや、レポートのトーンのリターンへのインパクトは一時的なものではないことを確認した。

このようにアナリストレポートの感情スコアが、株式リターンに対して有意に影響を与える研究報告には、次のような背景が考えられている。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られれば、高いリターンの獲得が可能となる。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と考えられている。

本稿では、幾つかの感情分析の方法により感情スコアを求めて、それらのスコアが将来のリターンと関係を捉える。将来のリターンとの関係について、幾つかの感情分析の方法のそれぞれで、どの程度、関係の強さが異なるかを捉えるものである。そして運用実務に応用する場合にどのようなしたら良いのかの考察を行う。

本稿で取り上げる感情分析の方法はスタンダードなものとする．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，アナリストレポートの文書にどの程度の予測力があるのかを直観的に捉えることができる．近年は Devlin, Chang, Lee and Toutanova (2019) の BERT (Bidirectional Encoder Representations from Transformers) などの Transformer 型の感情分析の実用化も見られる．Transformer 型とは直観的には，出力と入力を Attention というモデルのみで結ぶものである．その Attention とは，文中の単語の意味の理解にはどの単語に注目すればいいのかを表すニューラルネットワークである．日本語の BERT を用いた経済テキストデータの感情分析は青島・中川 (2019) などもある．BERT の利用に関しては，Wikipedia などコーパスとした学習を行いモデル化する方法や，学習済みモデルに調整を加える方法があるが，アナリストレポートなど表現が専門的な文書を対象にする場合には適用が難しいとの指摘も見られることから，今回は伝統的な感情分析の手法のみを対象とする．

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

本稿はアナリ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方法やデータを示す．

本稿で分析の対象とするアナリストレポートに関してはデータ取得が可能な 14 社とする．感情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でボリュームが様々である．また企業に関する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．しかしリターンの予測を目的とする感情スコアの算出には，レポートの内容全てを扱うのでは，そぐわない面もあると考えられるため，要約の 1 ページ目のみを用いる．

分析対象は TOPIX core30 対象銘柄とする．TOPIX core30 とは TOPIX ニューインデックスシリーズを構成する 8 指数のうちの 1 つで，TOPIX (東証株価指数) 算出対象銘柄の中で，日本を代表する大型株 30 銘柄で構成されるものである⁸．

学習データに関しては 2017 年 7 月～2019 年 8 月の期間に発行された 3565 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 9 月～2020 年 12 月までに発行されたレポートの 2266 本の感情スコアを算出する．

モデル化する上で学習サンプルサイズ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有用性があるのかも検討する。そして、それぞれのレポートに関して次に示す方法を用いて、将来の個別銘柄のリターンとの関係を検証する。

個別銘柄のリターンとの関係についての検証は、3 分位 (Tertile) ポートフォリオの手法を用いて、運用実務に適用する場合にどの程度の有用性があるのかを捉える観点からの分析を行う。3 分位とした理由は、①感情スコアは「ポジティブ」「ニュートラル」「ネガティブ」と 3 つの分類で捉えることが自然であることと、②Core30 を対象とした分析であることから、分位数を増やすと、それぞれに分位の銘柄数が少なくなり、検証の際に感情スコアとは関係ない個別銘柄要因の影響を受け易いからである。

テスト期間に関してそれぞれのレポートに感情スコアが付される。感情スコアの算出は次節で取り上げるが、レポート毎の感情スコアを TOPIX core30 対象の個別銘柄毎に月次で平均値を求める。

その平均値に関して、更に時系列方向に (1) 式の線形加重ウェイトを用いて、毎月の検証に使う感情スコアとする。本稿では感情ファクターと呼ぶことにする。

銘柄 i の当月の感情ファクター (検証に使う感情スコア)

$$\begin{aligned} &= \text{銘柄 } i \text{ の当月の感情スコア平均値} \times 4/10 \\ &+ \text{銘柄 } i \text{ の 1 月前の感情スコア平均値} \times 3/10 \\ &+ \text{銘柄 } i \text{ の 2 月前の感情スコア平均値} \times 2/10 \\ &+ \text{銘柄 } i \text{ の 3 月前の感情スコア平均値} \times 1/10 \end{aligned} \quad (1)$$

実際の検証は、毎月末で分析対象の Core30 について、(1) 式で求めた感情ファクターを基準にクロスセクションに昇順に銘柄数が等しくなるように 3 分位する。そして、各分位に該当する銘柄の等金額投資ポートフォリオの翌月のリターンを観察する。すなわち月次サイクルの 3 分位ポートフォリオの検証である。感情ファクターが最も大きい (ポジティブ) 第 3 分位のリターンがプラスの傾向となるのか、また、最も小さい (ネガティブ) 第 1 分位のリターンがマイナスの傾向となるのかを観察する。

検証期間に関しては次のようである。テスト期間の感情スコアは 2019 年 9 月～2020 年 12 月までであるが、(1) 式による感情ファクターの算出には 4 カ月が必要となる。このためテスト期間の感情ファクターは、2019 年 12 月～2020 年 12 月までとなり、これに対応したリターンを使った検証期間は 2020 年 1 月～2021 年 1 月までの 13 カ月となる。

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。

- (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

感情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⁹を用いる。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 12 月～2020 年 12 月までの 13 カ月間に作成されたレポートの評価を行う。

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 (2) 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。

$$\text{アナリストレポートの感情スコア} = \frac{\sum_{k=1}^n F_k \cdot \text{Weight}_k}{\sum_{k=1}^n F_k} \quad (2)$$

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アナリストレポート内に出現した頻度¹⁰、 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。

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

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。

まず、学習データの 2017 年 7 月～2019 年 8 月 (3565 本) とテストデータの 2019 年 9 月～2020 年 12 月 (2266 本) に発行された合計 5831 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 5831 本のレポートにおいて、先に抽出した Word である。従って、それぞれのレポートについて、8412 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 8412 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 3565 本の TF-IDF ベクトルを、レポート発行日の翌日から 5 日間でレポートの対象としている銘柄の対 Core30 指数に対するリターンがプラス (上昇) となったサンプルのベクトルと、マイナス (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 Core30 指数に対する超過リターンが上昇したか、或いは下落したかを基準に分類する。

次にテスト期間における 2266 本のレポートに対応した同数の TF-IDF ベクトルから、それぞれのレポートの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 3565 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、学習期間で上昇したサンプルのレポートに対応する TF-IDF ベクトルとのコサイン類似度の平均値から、同様に下落したサンプルのレポートの TF-IDF ベクトルと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 2017 年 7 月～2019 年 8 月 (3565 本) とテストデータの 2019 年 9 月～2020 年 12 月 (2266 本) に発行された合計 5831 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。ここまでは前述までの 2 つの方法と同じである。

次に、合計 5831 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 3565 本の分散表現のベクトルを、レポート発行日の翌日から 5 日間でレポートの対象としている銘柄の対 Core30 指数に対するリターンがプラス (上昇) となったサンプルのベクトルと、マイナス (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 Core30 指数に対する超過リターンが上昇したか、或いは下落したかを基準とする。そして上昇した分散表現のベクトルと、下落した分散表現のベクトルに分類する。

次にテスト期間 (2019 年 9 月～2020 年 12 月) における 2266 本の分散表現のベクトルから、その後の上昇、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 3565 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コサイン類似度の平均値から、下落サンプル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 2017 年 7 月～2019 年 8 月 (3565 本) とテストデータの 2019 年 9 月～2020 年 12 月 (2266 本) に発行された合計 5831 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、これまで紹介した方法と同様である。

次に、学習データの 3565 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 Word 2 Vec で求めたものを用いる¹¹。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、レポート発行日の翌日から 5 日間でレポートの対象としている銘柄の対 Core30 指数に対するリターンがプラス (上昇) となったサンプルのベクトルと、マイナス (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 Core30 指数に対する超過リターンが上昇したか、或いは下落したかを基準とする。そして上昇した分散表現のベクトルと、下落した分散表現のベクトルに分類する。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2017年7月～2019年8月(3565本)とテストデータの2019年9月～2020年12月(2266本)に発行された合計5831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5831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8Word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3565本の分散表現のベクトルを、レポート発行日の翌日から5日間でレポートの対象としている銘柄の対Core30指数に対するリターンがプラス(上昇)となったサンプルのベクトルと、マイナス(下落)したサンプルのベクトルの2つに分類する。この結果、上昇した分散表現のベクトルと下落した分散表現のベクトルに分類される。

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、下落した分散表現のベクトル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2266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9月～2020年12月(2266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)。

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

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	0.26	1.00			
(3)W2V	0.14	0.44	1.00		
(4)LSTM	-0.03	-0.04	0.03	1.00	
(5)D2V	0.24	0.44	0.16	0.01	1.00

(注) 2019 年 9 月～2020 年 12 月 (2265 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) セルサイドアナリ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、(4) LSTM は (1) 極性値、(2) TF-IDF との間で相関係数がそれぞれ、-0.03、-0.04 とわずかながらとは言え負値となった。手法間で傾向が異なるということは、感情分析を投資戦略に応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:0.44)。TF-IDF の手法は Bag of Words が基本である。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。

特に、(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) Word2Vec との 0.03 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を基準に算出した月次サイクルの個別銘柄別の感情ファクターとリターンとの関係を3分位ポートフォリオの手法で観察する。

検証に用いる感情ファクターは、2019年12月～2020年12月までの月次ベースで算出される。これに対応して2020年1月～2021年1月までの13カ月の期間で検証に用いる個別銘柄のリターンを用いる。

結果は表2に示される。T1、T2とT3の表中の「値」は、それぞれの分位に該当する銘柄の等金額投資ポートフォリオの月次リターンから、対象銘柄(Core30)の等金額投資ポートフォリオの差(超過リターン)を平均して年率換算したものである。またスプレッドはT3ポートフォリオのリターンからT1ポートフォリオのリターンを引いた値である。この値がプラスに大きいことは、ポジティブな感情ファクターが付された銘柄の将来リターンがプラスとなり、ネガティブな感情ファクターの銘柄の将来リターンがマイナスとなり銘柄選択効果が高いことを示す。

表2 感情ファクターの検証結果

	T1(低)		T2		T3(高)		スプレッド(T3-T1)	
	値	t値	値	t値	値	t値	値	t値
(1)極性値	-12.5%	-1.78	0.9%	0.18	12.2%	1.94 *	24.7%	1.98 *
(2)TF-IDF	-6.0%	-0.95	-7.7%	-1.11	14.5%	1.68	20.6%	1.51
(3)W2V	-10.5%	-1.44	3.8%	0.43	8.0%	0.96	18.4%	1.43
(4)LSTM	1.7%	0.23	-7.9%	-1.35	5.5%	1.44	3.8%	0.37
(5)D2V	-12.0%	-1.40	3.0%	0.35	9.7%	0.81	21.7%	1.14

(注1) 2020年1月～2021年1月までの13カ月間の検証。

(注2) 3分位ポートフォリオの月次サイクル検証。T1は感情ファクターが最も低い(ネガティブサイド)のポートフォリオ、T3は感情ファクターが最も高い(ポジティブサイド)のポートフォリオ。

(注3) T1、T2、T3の表中の「値」は、それぞれの分位に該当する銘柄の等金額投資ポートフォリオ月次リターンから、対象銘柄(Core30)の等金額投資ポートフォリオの差(超過リターン)を平均して年率換算したもの。

(注4) スプレッドはT3ポートフォリオのリターンからT1ポートフォリオのリターンを引いた値。

(注5) t値は、「値(平均値)」が0から有意に離れているかの値。*はt値が両側10%で有意に0から離れている。

(出所) セルサイドアナリストレポート、東京証券取引所のデータをもとに筆者作成

(4) LSTMを除いて他の4手法はT1の値がマイナス、T3がプラスの傾向となり、T3-T1のスプレッドがプラスの傾向を見せている。これは、ポジティブ感情ファクターの将来のリターンがプラス、ネガティブ感情ファクターの将来のリターンがマイナスとなり、感情ファクターが銘柄選択にある程度、有効に機能する可能性を示す。以下では個別の感情ファクター毎に傾向を観察する。

(1) 極性値を使った感情ファクターの銘柄選択効果は最も有効性が高かった。スプレッド(リターン)の(平均)値は24.7%と高く、10%有意となっている。アナリ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シンプルな方法に効果があった。

Bag of Words をベースとする (2) TF-IDF の手法も一定の銘柄選択効果が見られた。統計的に有意とは言えないものの、スプレッドの値は 20.6% と高かった。基本的にポジティブとネガティブの Word の集計を基本とするシンプルな方法に効果が高かったと見られる。

一方、(3) Word2Vec (W2V) の手法は、前述した 2 つの手法と比べると銘柄選択効果がやや低下しているものの、スプレッドの値は 18.4% と将来リターンとの関係がある程度は見られる。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。レポートに表れるアナリストの特徴的なクセなどの特徴が把握できることが将来のリターン予測に有効となるかもしれない。

(4) LSTM はスプレッドが 3.8% と低く、将来リターンとの関係が殆どみられなかった。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 の手法を更に発展させる形である。ただ、今回はアナリストレポートの表紙部分のみを用いた感情分析であったことから、よりシンプルにレポートの感情を捉える手法が将来のリターンとの関係では効果的なものかもしれない。

しかしながら 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) は予測力の観点で一定の効果が確認される。スプレッドが 21.7% と今回取り上げた手法のなかで 2 番目の水準となった。Word2Vec と同様にレポートに表れるアナリストの特徴的なクセなどの特徴が把握できることが、ある程度の銘柄選択効果につながったと考えられる。

これらの分析結果に関しては注意点もある。(1) 極性値を使った感情ファクター以外は、将来のリターンとは統計的に有意な関係とはならなかったからである。しかし、一定の効果が確認されたことは実務面での有用性の観点から評価ができる。

5 おわりに

本稿では、アナリ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を、月次サイクルで集計した個別銘柄の感情ファクターに変換して、将来の個別株式リターンとの関係を分析した。

分析結果から、(4) LSTM を除いた 4 手法は、感情ファクターが銘柄選択にある程度、有効に機能する傾向が見られた。なかでも (1) 極性値を使ったシンプルな手法で算出した感情ファクターの銘柄選択効果が最も高かった。今回はアナリストレポートの表紙部分のみを用いた感情分析であった。表紙部分はレポート全体の要約の部分であるため、読者に分かり易く伝えるために完結な表現が多く使われる傾向がある。このため複雑なモデルと比べてシンプルな (1) 極性値の集計が、将来のリターンとの関係でより妥当な感情スコアが求められたのかもしれない。

しかしながら、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。レポートの表紙部分の Word の前後関係に関するアナリストの表現のクセなどの特徴を捉えられると、銘柄選択に有効となる可能性がある。

本稿はアナリストレポートを対象としたものだが、同様の 5 つの感情スコアの算出方法を用いて、ストラテジストレポートの感情スコアと将来のリターンとの関係を検証した吉野 (2021) では、今回の分析とは対照的な結果となっていた。吉野 (2021) では、(1) 極性値で求めた感情スコアは将来のリターンとの関係が見られず、将来の株式リターンの予測は難しかった。このようにアナリストレポートとストラテジストレポートでは、将来のリターンと強い関係が見られた感情スコア算出方法が異なっていた理由には、アナリストレポートとストラテジストレポートでは感情スコアの趣旨が異なるという点があるからだろう。アナリストレポートに関して言えば、アナリストが業績予想を修正する前に、レポートの文書から業績予想の修正の予兆を捉えられれば、高いリターンの獲得が可能となる。感情スコアは、こうした予兆と関係するものと考えられる。一方、ストラテジストレポートに関しては、ストラテジストが相場水準の見通しを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。しかしストラテジストは市場の専門家であり、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可能性があるため将来のリターンと関係が見られると考えられる。このため、シンプルな (1) 極性値からは将来のリターンの予測力が小さく、Word の並び順などの類似性を捉える (4) LSTM のリターンの予測力が高かったと考えられる。

本稿の分析から手法間で感情スコアの傾向が大きく異なることや、将来のリターンとの有効性も異なることから、実務に利用する際には手法の選択が重要であることが分かった。実際に実務に応用する上では、各感情スコアの特徴を捉えたうえで、アンサンブル型の利用も効果的となるかもしれない。

今回の分析では学習期間においてレポートをポジティブとネガティブに分類するために、5 日間の株式リターンの騰落を用いた。分類の方法には、レーティングの変更などが使われるケースもあり、様々な分類方法も検討する必要があるだろう。

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 森田充教授より有益なコメントを頂いた。ここに記し、感謝申し上げる。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。

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 株式を売る側。証券会社や投資銀行などを指す。
- 2 企業の財務分析、業界分析、経営者インタビューなどをもとに、株式の投資価値の分析・評価をおこなう。本稿ではセルサイドの企業アナリストが作成する企業分析レポート対象に分析を行う。セルサイドの企業アナリストは証券会社の調査部などに所属しており、彼らの行う企業の分析・評価はレポートを通じて投資家に提供される。
- 3 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。
- 4 「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。
- 5 Apple の Siri や Amazon の Alexa などが代表的なもの。
- 6 「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。
- 7 一般にはアナリストレポートを提供する情報ベンダーやセルサイドとビジネスで関係する機関投資家などに限られてしまうため。
- 8 TOPIX core30 の詳細は日本証券取引所のウェブサイト <https://www.jpx.co.jp/>参照。
- 9 <https://sites.google.com/socsim.org/izumi-lab/tools/>金融専門極性辞書
- 10 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。
- 11 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。しかしアナリ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。

（参考文献）

- 青島智久・中川慧（2019）,「日本語 BERT モデルを用いた経済テキストデータのセンチメント分析」『2019 年度人工知能学会全国大会論文集第 33 回』.
- 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.
- 平松賢士・三輪宏太郎・酒井浩之・坂地泰紀（2021）,「アナリストレポートのトーンの情報価値」『証券アナリストジャーナル』 59（2）, 86-97.
- 吉野貴晶（2021）,「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」『国際マネジメント研究』 10, 17-31.
- 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.
- Huang A., A. Zang and R. Zheng (2014), “Evidence on the Information Content of Text in Analyst Reports,” *The Accounting Review* 89(6) , 2151–2180.
- Le, Q. , and T. Mikolov (2014), ”Distributed Representations of Sentences and Documents”, *CoRR*, abs/1405. 4053, 1-9.
- Miwa, K. (2021), ”Levels versus Changes: Information Contents of Textual Information,” *International Review of Finance*, In press, 1-8.
- 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.
- Twedt, B. and L. Rees (2012), “Reading between the Lines: An Empirical Examination of Qualitative Attributes of Financial Analysts’ Reports,” *Journal of Accounting and Public Policy*, 31, 1-12.