

Verification of Predictive Power of Quarterly Financial Statement Text Information for Stock Price Fluctuations

Mana MATSUDA* and Hiroshi TSUDA*

(Received October 3, 2023)

Due to the development of natural language processing technology, there has been a lot of research on text mining and its use for investment decisions. In this study, we quantified the quarterly financial statement text information using Sentence-BERT and verified its explanatory power for stock price fluctuations using regression trees. We also used classification trees to examine the explanatory power for the direction of stock price fluctuations. We then conducted an investment simulation and confirmed that excess returns were generated relative to the Nikkei 225, suggesting the predictive power of the quarterly financial statement text information on stock price fluctuations.

Keywords: natural language processing, stock price fluctuation, decision tree, deep learning

キーワード: 自然言語処理, 株価変動, 決定木, 深層学習

決算短信テキスト情報の株価変動に対する予測力の検証

松田 眞, 津田 博史

1. はじめに

近年, 自然言語処理技術の発展に伴って, 深層学習などを用いてテキスト情報の定量化やテキストマイニングを行い, 投資判断に活用しようとする研究が盛んである. これらの研究では, SNS の投稿やニュース記事, アナリストレポートなど様々なテキスト情報が用いられる. 本研究で用いる決算短信もその1つである.

例えば, 北森ら(2017)¹⁾や酒井ら(2019)²⁾, 河村ら(2021)³⁾は, 決算短信から業績予測文や業績要因文の抽出を行った. これらの研究は, 個人投資家の投資判断を支援することが目的である. 特に,

河村らは, 事前学習済みBERT⁴⁾を用いて, 決算短信に含まれる文が「業績要因を含む業績予測文」であるかどうかを判定する分類モデルを学習した. ここで, 学習用データである「業績要因を含む業績予測文」は自動生成されている. 具体的には, 業績予測文の文末手がかり表現や業績要因の手がかり表現, 企業キーワードをもとに「業績要因を含む業績予測文」の正例と負例の条件を設定し, 自動生成を行った. その結果, 「業績要因を含む業績予測文」に対する高い分類性能が得られた. この研究は, テキストマイニングによって, 決算短信から投資判断を行ううえで重要な文章を抽出することが目的であった.

*Department of Mathematical Science, Doshisha University, Kyoto

Telephone : +81-774-65-6681, E-mail : ctwh0908@mail4.doshisha.ac.jp, htsuda@mail4.doshisha.ac.jp

一方で、決算短信のテキスト情報を定量化することで、金融市場との関係を明らかにしようとした研究も存在する。はじめに、決算短信を分析対象とした白方(2018)⁹⁾と山本ら(2020)⁶⁾について述べる。白方は決算短信テキスト情報に対して、企業業績に対する極性値を付与し、業績予想の売上達成率と極性値の関係を調べた。その結果、極性値が大きいほど業績予想の売上達成率が大きいことが分かり、極性値が決算短信テキスト情報を反映した値であることを示唆した。また、極性値の範囲で分類された銘柄で構成される2つのポートフォリオに対して、リターン之差を検証するスプレッドリターン検証を行い、極性値と株式リターンの関係も調べたが、企業業績とは異なり、極性値が大きいほど株式リターンが大きいという関係は一貫しておらず、提案手法で算出した極性値だけでは株価予測が難しいことを示唆した。一方で、山本らは、決算短信に含まれる単語に肯定語・否定語スコアを付与し、それらを集計することで決算短信のテキストスコアを算出した。さらに、テキストスコアを既存の投資戦略で用いられる業績数値情報で回帰した残差を決算短信の投資スコアとした。山本らは、白方と同様、スプレッドリターン検証を行い、投資スコアが正に有意なスプレッドリターンをもつことを示した。つまり、決算短信テキスト情報が既存の投資情報にはない効果をもつことを示唆した。次に、SNSの投稿で暗号通貨の価格変動に関するトレンドを予測する効果を調べた関岡ら(2023)⁷⁾について述べる。関岡らはSentence-BERT⁸⁾(以後、SBERT)を用いてSNSの投稿のテキスト情報を定量化し、このテキスト情報の特徴量と出来高や取引量などの数値情報の特徴量を用いて、決定木をベースにした機械学習モデルであるLightGBM⁹⁾を学習し、トレンド(急降下、急上昇、急激な変動なし)の分類を行った。そして、数値情報の特徴量だけの場合と比較することで、トレンド予測においてSNSの投稿のテキスト情報を用いる効果を調べた。その結果、SNSのテキスト情報を用いることは急激な価格変動を予測するうえで役立つことが分かり、出来高などの

数値情報と合わせて予測に用いることで、さらに効果を発揮することを示唆した。

以上の先行研究をふまえて、決算短信テキスト情報の定量化には企業業績に対する極性値が用いられることが多い。そこで、松田・津田(2023)^{10,11)}では、決算短信テキスト情報を、SBERTを用いて定量化し、テキスト情報の特徴量をもつ株価変動に対する説明力を検証した。具体的には、決算短信に含まれる文をSBERTでベクトル(分散表現, 384次元)に変換する。つまり、ここでは決算短信に含まれる文の数だけベクトルが獲得できる。しかし、決算短信ごとに含まれる文の数は異なるため、松田・津田(2023)^{10,11)}は、分散表現の成分ごとにヒストグラムを作成し、相対度数を重みとした加重平均を算出することで、この加重平均を要素にもつ384次元のベクトルを作成した。こうすることで、決算短信に1つのベクトルを対応付けることができ、松田・津田(2023)^{10,11)}は、このベクトルの各成分を説明変数として、株価変動に対する説明力を企業ごとに検証し、高い説明力をもつ成分について考察した。ここで、株価変動として表すのは、固有銘柄*i*の四半期収益率 r_{i,t_j} を日経平均の四半期収益率 r_{k,t_j} で回帰したときの残差 ε_i である。(決算短信発表日間の四半期)

$$r_{i,t_j} = a + br_{k,t_j} + \varepsilon_i \quad (j = 0, 1, 2, \dots) \quad (1)$$

この残差は、個別銘柄の株価変動のうち、日経平均で説明可能な市場要因による株価変動の影響を除いた、固有要因による株価変動を表し、この残差が正であるとき、その銘柄は日経平均に対して残差リターンをもつことになる。このことから、残差の予測精度を高めることで、日経平均に対して超過リターンを生むポートフォリオを作成するため、残差を株価変動として用いる。また、決算短信テキスト情報の株価変動に対する説明力を検証するための回帰モデルの作成には、松田・津田(2023)¹⁰⁾では回帰分析を、松田・津田(2023)¹¹⁾では回帰木を用いた。その結果、回帰分析を用いた場合と比較して、回帰木の方が、多くの企業で株価変動に対してより高い説明力をもつテキスト情報の特徴量(成分)を見出すことができた。これ

は、回帰木では、決算短信テキスト情報の特徴量と株価変動の間の非線形なパターン、ひいては SBERT の分散表現の成分と株価変動の間の非線形なパターンを認識できたからだと考える。

本研究では松田・津田(2023)^{10,11)}において、決算短信テキスト情報の特徴量として、SBERT の分散表現の成分ごとに加重平均を算出していたのに対し、平均、標準偏差、歪度を算出し、それらの特徴量に用いた。また、株価変動として残差を対象にした目的である、日経平均に対する超過リターンの有無を確かめるために、投資シミュレーションを行った。これによって、本研究の決算短信テキスト情報の特徴の株価変動に対する予測力を検証する。加えて、松田・津田(2023)^{10,11)}と同じテキスト情報の特徴量を用いて、株価変動の正負を分類する分類木を作成し、株価変動の方向に対する説明力を検証する。

本稿では、第 2 章で使用データについて、第 3 章で研究手法について、第 4 章で分析結果について、第 5 章でまとめ・今後の課題について述べる。

2. 使用データ

本研究では、投資シミュレーションをふまえて、2016 年 1 月 1 日から 2022 年 12 月 31 日までの期間に、日経平均構成銘柄であった企業を対象とする。つまり、構成銘柄の定期入れ替えや臨時入れ替えて除外された企業も分析対象とする。ただし、{銀行、証券、保険、その他金融}に属する企業を除き、3 月

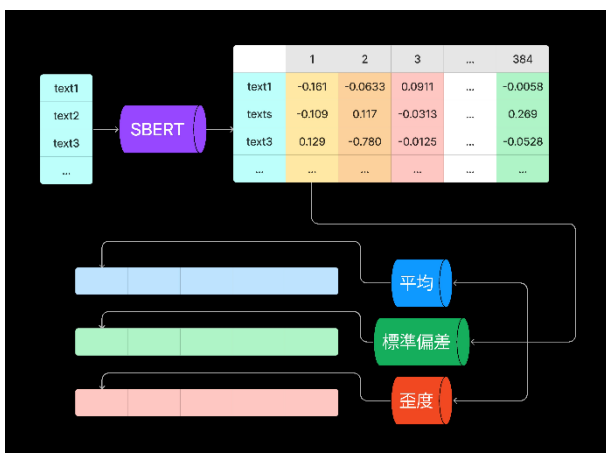


Fig. 1. Flowchart of feature acquisition.

決算の企業に限った 29 業種 184 社を分析対象とする。これらの企業の 2012 年 1 月 1 日から 2022 年 12 月 31 日までに発表された決算短信および、決算短信発表日(決算短信発表日が営業日でない場合、その前営業日)の終値、同日の日経平均の終値などを使用した。

3. 分析手法

3.1. 決算短信テキスト情報の特徴量の算出

本研究では、松田・津田(2023)¹⁰⁾と同様の方法で、pdf 形式の決算短信からテキスト情報を抽出しているため、抽出方法の詳細は松田・津田(2023)¹⁰⁾を参照せよ。本節では、決算短信に含まれる複数の文から、決算短信テキスト情報の特徴量を算出する方法を中心に述べ、最後に説明力の検証方法について述べる。Figure 1 では、初めに、決算短信から抽出した複数の文を SBERT に入力している。これによって、各文に対応する 384 次元の分散表現が獲得できる。その後、これらのベクトルを統合して、3 つの特徴量を算出する。Figure 1 では各文に対応する分

平均	1	2	3	...	384
2012/01/26	0.00355	-0.0575	-0.00233	...	0.0185
2012/04/26	-0.0244	-0.0481	-0.0729	...	0.0364
...
2022/07/30	-0.0185	-0.0571	-0.012	...	0.0158

標準偏差	1	2	3	...	384
2012/01/26	0.133	0.150	0.107	...	0.118
2012/04/26	0.140	0.145	0.124	...	0.109
...
2022/07/30	0.142	0.149	0.108	...	0.119

歪度	1	2	3	...	384
2012/01/26	0.347	-0.268	-0.612	...	0.888
2012/04/26	0.468	-0.402	-0.0724	...	0.0316
...
2022/07/30	0.105	-0.460	-0.247	...	0.432

Fig. 2. Databases of text information features.

散表現の第1成分の処理を示している。具体的には、第1成分のデータに対して、{平均, 標準偏差, 歪度}をそれぞれ算出し、対応する特徴量の第1成分とする。この処理を各成分に対して行うことで、{平均, 標準偏差, 歪度}に対応する384次元の特徴量が3つ作成できる。例として、Fig. 2には、Fig. 1のフローによって、ある企業の分析対象期間に発表された各時点の決算短信を定量化した結果を示す。それぞれ、SBERTの分散表現に対して{平均, 標準偏差, 歪度}を算出した結果である。ここで、回帰木の説明変数には、Fig. 2に示す{平均, 標準偏差, 歪度}に対応する3つのデータベースから複数(または1つ)を選択し、各成分の特徴量を用いた。目的変数には株価変動(残差)を用いる。そのため、データベースの組み合わせに対して、1社につき384個の回帰モデルが作成される。ここで、決定木は、学習用データ(7割)で各ノードにおける分割条件を決定木に学習させ、検証用データ(3割)に対する決定係数で精度を評価した。この決定係数が最大になる木の深さを最適な深さ(初期値は2)として選択した。最後に384個の回帰モデルのうち、決定係数が最も高いモデルを選択し、その決定係数をもって、決算短信テキスト情報の株価変動に対する説明力を評価する。

3.2. 分類木を用いた株価変動の方向の分類

はじめに、分類木とは、分岐によってデータの分類を行う決定木である。本研究では、説明変数に、松田・津田(2023)^{10,11)}と同じテキスト情報の特徴量を用いる。つまり、SBERTの分散表現の各成分のデータに対して加重平均を算出することで得られる各

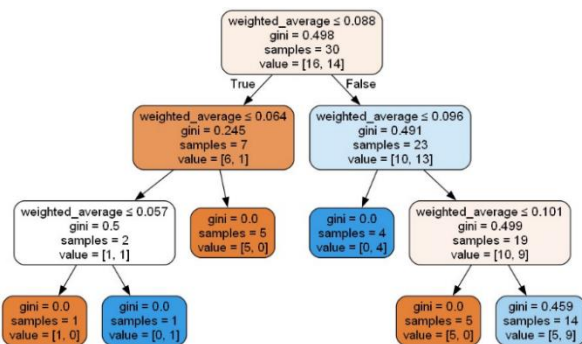


Fig. 3. Example of decision tree.

成分の特徴量である。また、目的変数は、株価変動

の符号によって{-1, 0, +1}を対応させたカテゴリカル変数である。Figure 3のように、回帰木と同様に、分類木においても、各ノードにおいて説明変数を用いた分割条件を学習する。そして葉ノードに属するデータの目的変数の値で多数決を行った結果が、分類木の予測である。したがって、Fig. 3の葉ノードのうち右端のノードに着目すると、 $value = [5, 9]$ とあるからこのノードに属するデータには、目的変数の値が-1であるデータが5個、1であるデータが9個あることから、多数決により1が予測値になる。また、分類木ではジニ不純度という評価指標を用いて最適な分割条件を決める。ジニ不純度は以下の式で定義される。

$$I(t) = 1 - \sum_{l=1}^C P(l|t)^2 \quad (2)$$

C はカテゴリ数であるから、今回は3(株価変動が0のデータがない場合は2)である。 $P(l|t)$ はあるノード t に属するデータの内、カテゴリが l であるデータの割合を表す。したがってジニ不純度は、あるノードのデータが全て一方のカテゴリに属する場合0になるような指標である。さらに、分岐によるジニ不純度の変化を表す指標を情報利得といい、次の式で定義される。

$$IG(D) = I(D) - \frac{N_{yes}}{N_d} I(D_{yes}) - \frac{N_{no}}{N_d} I(D_{no}) \quad (3)$$

ここで、 N_d は分岐前のノードに属するデータ D の数で、 N_{yes} は条件式を満たす場合の分岐先のノードのデータ D_{yes} の数、 N_{no} は条件式を満たさない場合の分岐先のノードのデータ D_{no} の数を表す。つまり、分岐前のノードのデータのジニ不純度 $I(D)$ からデータ数で調整した分割後の2つのノードのデータのジニ不純度 $\frac{N_{yes}}{N_d} I(D_{yes})$, $\frac{N_{no}}{N_d} I(D_{no})$ の合計を引くこと

で、分岐によるジニ不純度の変化を計算している。よって、分類木では条件利得が最大になる分割条件を各ノードで学習する。分類木の深さについても3.1節で述べた方法で決める。4章では分類木を用いて、決算短信テキスト情報がもつ株価変動の方向に対する説明力を検証した結果について述べる。

3.3. 投資シミュレーション

本研究では、決算短信テキスト情報の株価変動に対する説明力を検証する。ここで、本研究における株価変動は、固有銘柄の日経平均に対する残差リターンを表す。本研究では、3.1 節で述べた説明変数に対して回帰木を用いる方法や、3.2 節で述べた通り分類木を用いる方法で、分析対象期間の株価変動に対する決算短信テキスト情報の説明力を評価する。投資シミュレーションの目的は、本研究で獲得したテキスト情報の特徴量がもつ株価変動に対する予測力を評価することである。投資シミュレーションは次のルールで行う。

1. 日経平均構成銘柄で構成されるポートフォリオを作成する。

(ア) 定期入れ替えや臨時入れ替えは日経平均と同様に行う。(新規上場銘柄を追加する場合を除く)

2. ポートフォリオの指数について

(ア) 日経平均は 1950 年 9 月 7 日の東証第一部修正平均株価が起源であり、当初の算出方法は構成銘柄の株価合計を銘柄数で割った単純平均であった。つまり、Eqs. (4)-(5)において、株価換算係数を 1、除数を 225 として求まる値と等しい。その後、指数の連続性を維持する目的で、Eqs. (4)-(6)に従って株価換算係数や除数を調整する方法に変更された。Eq. (6)の翌日構成銘柄の基準株価合計は、各銘柄の当日の終値と翌日の株価換算係数の積の合計である。また、翌日の株価換算係数は、銘柄入れ替えが生じた場合には、基準日(1 月末または 7 月末)において追加銘柄の採用株価が構成銘柄の採用株価合計の 1%を上回らない最大の値(0.1 以上 0.9 以下、0.1 刻み)として決まる。ただし、銘柄入れ替えや株式分割(併合)が生じない場合は、当日の株価換算係数を引き継ぐ。Equations (4-5)より日経平均の算出式は Eq. (7)と書き換えられる。各銘柄の株価換算係数を除数で除した値(以後、調整後株価換算係数)を

株数と捉えると、日経平均は当日の終値で換算した値と捉えられる。また、日経平均では銘柄入れ替えによる株価換算係数の更新に伴い、除数が更新されるが、Eq. (6)による除数の更新から、銘柄入れ替えによって日経平均は変動しないことが分かる。これは、日経平均が構成銘柄の市場要因による株価変動の影響だけを受けて変動する性質を満たすためである。これに対して、ポートフォリオの指数は、シミュレーション期間の初日の日経平均の調整後株価換算係数を初期値に用いて、銘柄入れ替えに加えて決算短信の発表に伴って調整後株価係数を更新する。ここで指数の算出式は日経平均と同様 Eq. (7)を用いて、リバランスで指数の変動が生じないように調整後株価係数を更新する。したがって、ポートフォリオ指数は、構成銘柄の市場要因による株価変動と、決算短信のテキストから推定した各銘柄の株価変動の予測の影響を受けて変動する。

$$\text{日経平均} = \frac{\text{構成銘柄の採用株価合計}}{\text{当日の除数}} \quad (4)$$

$$\text{採用株価} = \text{当日の終値} \times \text{当日の株価換算係数} \quad (5)$$

$$\text{翌日の除数} =$$

$$\text{当日の除数} \times \frac{\text{翌日構成銘柄の基準株価合計}}{\text{当日構成銘柄の採用株価合計}} =$$

$$\text{当日の除数} \times$$

$$\frac{\sum_{\text{翌日の構成銘柄}} \text{当日の終値} \times \text{翌日の株価換算係数}}{\sum_{\text{当日の構成銘柄}} \text{当日の終値} \times \text{当日の株価換算係数}} \quad (6)$$

$$\text{日経平均} = \sum_{i=1}^{225} \text{終値}_i \times \frac{\text{株価換算係数}_i}{\text{除数}} \quad (7)$$

3. リバランスについて

本研究においてリバランスとは、調整後株価換算係数を更新することを意味する。また、銘柄入れ替えに伴うリバランスは日経平均と同様の手法を用いるため、ここでは決算短信発表に伴うリバランスに限定して述べる。

- (ア) 実際に決算短信の内容を考慮して、株式を売買できるのは発表日の翌営業日であることから、決算短信発表日の翌営業日にリバランスを行う。そのため、調整後株価換算係数の計算に用いる株価は、翌営業日の始値である。(ただし、銘柄入れ替えについては前もって追加(除外)される銘柄やその日付が知らされるため、日経平均と同様に当日の終値を用いる。)
- (イ) 過去に発表された収集可能な直近 20 個の決算短信とそれらに対応する株価変動のデータより回帰木を学習・選択する。(3.1 節で述べた手法と同様。)
- (ウ) 発表された決算短信から特徴量を取得し、(イ)の回帰木を用いて株価変動の予測値を取得する。
- (エ) (ウ)の予測値をもとに調整後株価換算係数を更新する。以下の通り更新を行うことで、前営業日に決算短信を発表した銘柄の調整後株価換算係数のみを決定木の予測値に従って更新でき、指数を変動させることなくリバランスできる。言い換えると、前営業日に決算短信が発表された銘柄の株式だけを始値で売買し、なおかつ売買によって運用資金が変動しないようにリバランスを行う。
- ① 前営業日に決算短信が発表された銘柄の株価変動に対する予測値 $\{p_i | i = 1, 2, \dots, N\}$ が得られているとする。予測値の大きさを $\min_i p_i - 0.1$ を基準に求める。
 - ② 銘柄 $i = 1, 2, \dots, N$ 以外の前営業日の調整後株価換算係数を用いて、当日に銘柄 $i = 1, 2, \dots, N$ に割り振られる運用資金 (N 銘柄の調整後株価換算係数 \times 始値の合計値)を求める。
 - ③ ①で求めた予測値の大きさの比率で②の運用資金を銘柄 $i = 1, 2, \dots, N$ に配分し、の各銘柄の運用資金

(調整後株価換算係数 \times 始値)を決める。それを各銘柄の始値で割ることで調整後株価換算係数を求める。

以上のルールに従って、2021 年 10 月 1 日から 1 年

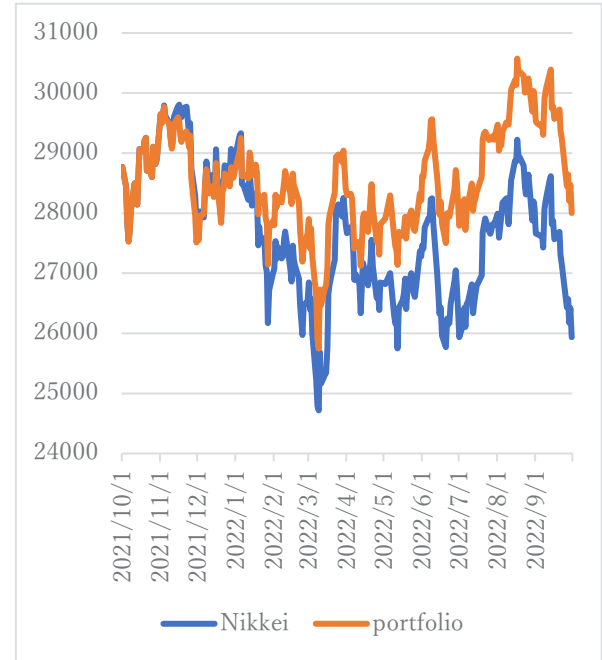


Fig. 4. Changes in index.

間の投資シミュレーションを行い、日経平均に対す

Table 1. Adjusted R-squared.

average	skewness
0.429	0.456
standard deviation	average, standard deviation
0.442	0.345
average, skewness	average, skewness, standard deviation
0.356	0.244

る超過リターンを検証する。

4. 結果・考察

4.1. 決定係数の比較

はじめに、3.1 節で述べた手法により、企業ごとに、分析対象期間の株価変動に対する決算短信テキスト情報の説明力を検証した結果について述べる。具体的には、本研究で用いたテキスト情報の特徴量(以後、平均、標準偏差、歪度の特徴量)と、松田・津田(2023)¹⁰⁾¹¹⁾で用いた特徴量(以後、加重平均の特徴量)による自由度調整済み決定係数の結果と比較する。Table 1 は各説明変数を用い

た場合の分析対象企業の自由度調整済み決定係数の平均値を示している。加重平均の特徴量では、決定係数の平均値が 0.458 であったことから、平均、標準偏差、歪度の特徴量のいずれの組み合わせも加重平均の特徴量の決定係数を上回ることができなかった。このことから、本研究で説明変数に用いた統計量の中で、加重平均が決算短信テキスト情報を表すうえで、最も有効であることが分かった。次に、分類木を用いて、加重平均の特徴量をもつ株価変動の方向に対する説明力を検証した結果を述べる。分析対象企業の決定係数の平均値は 0.91 である。この結果から、分類木は、テキスト情報の特徴量をもつ株価変動の方向に対する高い説明力を見出せることが分かる。

4.2. 投資シミュレーションの結果

本研究で株価変動に対する説明力を検証したテキスト情報の特徴量のうち、説明力が最大であったのは、加重平均の特徴量であった。このことから、加重平均の特徴量を用いて 3.1 節の方法で投資シミュレーションを行った。Figure 4 には、回帰木の株価変動に対する予測をもとにリバランスを行ったポートフォリオの指数(portfolio)と日経平均(Nikkei)の推移を示している。Figure 4 から分かるように、ポートフォリオは日経平均に対して多くの時点で超過リターンを生むことが分かる。このことから、加重平均の特徴量を用いて作成した回帰木は、株価変動の大きさ及び方向に対して予測力を有していることが示唆できたと考える。

5. まとめ

本研究では、29 業種 184 社を対象に、2012 年 1 月 1 日から 2022 年 12 月 31 日までに発表された決算短信を用いて分析を行った。具体的には、①SBERT の分散表現の各成分に対して{平均, 歪度, 標準偏差}を計算し, {歪度}, {標準偏差}, {平均, 歪度}, {平均, 標準偏差}, {平均, 歪度, 標準偏差}の組み合わせに対して, 各成分のデータを説明変数に用いて, 回帰木を用いて株価変動に対する説明力を検証した。また, ②松田・津田(2023)¹¹⁾ と同じテキスト

情報の特徴量を説明変数にして, 分類木を用いて株価変動の方向に対する説明力を検証した。①については, いずれの説明変数も松田・津田(2023)¹¹⁾ で用いたテキスト情報の特徴量をもつ株価変動に対する説明力より高い説明力を見出すことはできなかった。このことから本研究で用いた統計量と比較して, SBERT の分散表現の各成分に対する加重平均が, 決算短信テキスト情報を表現する特徴量として効果があることが示された。②については, 分類木を用いると, テキスト情報の特徴量をもつ株価変動の方向に対する高い説明力を見出せることが分かった。さらに, 決算短信テキスト情報の特徴量(加重平均の特徴量)で作成した回帰木を用いて株価変動を予測する性能を検証すべく, 投資シミュレーションを行って, 日経平均に対する超過リターンを確かめた。その結果, 回帰木の予測結果をもとにリバランスを行ったポートフォリオは, 日経平均に対して超過リターンを生むことが分かった。このことから, 回帰木は, 株価変動に対する予測力を有することを示唆できたと考える。以上をまとめると, 本研究で用いた統計量のうち, 加重平均を用いてテキスト情報の特徴量を作成することが, 説明力を見出すうえで効果を有していることが分かり, さらに投資シミュレーションの結果から, 回帰木を用いて, 決算短信テキスト情報の特徴量が将来の株価変動に対しても予測力を有していることを示唆できた。

6. 今後の課題

決算短信のテキスト情報を対象に, 株式市場との関係を研究した先行研究において, テキスト情報の定量化に業績に対する極性値が用いられることが多い。そこで, 本研究では, SBERT の分散表現を用いて特徴量を作成した。そして, 投資シミュレーションによって日経平均に対して超過リターンが生まれることを確認することで, この特徴量が有効性を示唆した。そこで, 本研究の決算短信テキスト情報の定量化手法を極性値に代替し, 同様の手法で超過リターン(残差リターンに対する説明力)を算出し, 本研究の超過リターン(説明力)と比較することで, 極

性値による定量化手法に対する、本研究の定量化手法の優位性を示すことが今後の課題である。さらに、本研究の定量化手法を実際の株式のポートフォリオ運用に応用するには、売買手数料、売買時のマーケットインパクト、キャッシュフローマネジメントなどを考慮する必要がある、ポートフォリオマネジメントの研究が必要である。

参考文献

- 1) 北森 詩織, 酒井 浩之, 坂地 泰紀, “決算短信 PDF からの業績予測文の抽出”, 電子情報通信学会論文誌 D, Vol.J100-D, **2**, 150-161, (2017).
- 2) 酒井 浩之, 松下 和暉, 北島 良三, “学習データの自動生成による決算短信からの業績要因文の抽出”, 日本知能情報ファジィ学会誌, **31**[2], 653-661, (2019).
- 3) 河村 康平, 高野 海斗, 酒井 浩之, “決算短信からの業績要因を含む業績予測文の抽出”, 人工知能学会全国大会論文集, **35**, (2021).
- 4) J.Devlin, M.Chang, K.Lee, and K.Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, *arXiv:1810.04805*, (2018).
- 5) 白方 健司, “自然言語処理と機械学習による株式市場の予測”, 同志社大学修士論文, (2018).
- 6) 山本 零, 川代 尚哉, 栗田 昌孝, “決算短信と四季報テキスト情報の投資戦略への利用可能性検証”, ジャプリー・ジャーナル, **18**, 46-62, (2020).
- 7) S.Sekioka, R.Hatano, and H.Nishiyama, “Market Prediction Using Machine Learning Based on Social Media Specific Features”, *Artificial Life and Robotics*, Springer Nature, (2023).
- 8) N.Reimers and I.Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, *arXiv:1908.10084*, (2019).
- 9) G.Ke, Q.Meng, T.Finley, T.Wang, W.Chen, W.Ma, Q.Ye, and T.Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”, *Adv Neural Info Process Syst* 30, 3146–3154, (2017).
- 10) 松田 眞, 津田 博史, “Sentence-BERT を用いた決算短信のテキスト情報の株価変動に対する統計的有意性の検証”, 同志社大学ハリス理化学研究報告, **64**[2], 37-48, (2023).
- 11) 松田 眞, 津田 博史, “決定木回帰を用いた決算短信テキスト情報の株価変動に対する説明力の検証”, 同志社大学ハリス理化学研究報告, **64**[3], 19-27, (2023).