

# 決算短信発表前後のインターネット掲示板への 投稿分析のための考察\*

富田 知嗣

はじめに

証券市場におけるディスクロージャー制度に基づく情報は、投資家であっても投資家でなくても利用可能であるが、既に証券投資を行っているか将来的にそうするかもしれないかは別にして、投資家が証券市場において証券投資に関する意思決定に資するための情報である。これらの情報の中には、有価証券報告書や決算短信で示される会計情報も含まれる。

会計情報の内容を解釈することについては、『財務会計の概念フレームワーク』で示されるとおり、「投資家の中には会計情報の分析能力に優れた者のほか、自らは十分な分析能力を持たず専門家の助けを必要とする者も含まれている」<sup>1)</sup>。特に個人投資家は、機関投資家と比較して、証券市場で投資を行うために必要とされる知識だけでなく関連する情報の収集能力や情報を解釈する能力にばらつきが大きいと考えられる。このような知識や能力において、劣後する投資家は、証券市場から退出するか適切な方法によって不足を補いながら証券市場にとどまることになる。適切と思われる方法についていくつかの方法が考えられるが、その一つとして自らでは解釈できない情報を解釈した情報もしくは自らが解釈するための助けとなる情報を入手する方法がある。このような情報を入手するためには、関連するコミュニティに参加する、有料の情報を購入する、無料で公開されている情報を探すなどが存在する。

決算日から発表まで相応の時間を要する有価証券報告書に比して、決算短信は速報的に迅速に開示されるため、決算短信の発表時には注目を集めやすい。結果として、決算短信によって開示される会計情報は投資家による企業価値の再評価もしくは修正を促進し、短時間のうちに新しい情報に基づく株価に修正すべく株価変動をもたらす。最近のICTの発展と普及はその速度を加速していると考えられる。個人投資家のうち情報解釈や分析能力を十分にもたない個人投資家が、短時間に生じる株価変動に対応するためには、決算短信発表による株価変動の影響を受けない投資方法を選択するか、解釈情報（解釈の手がかりとなる情報を含む）を入手して

\* 本研究は、2019年度関西大学研修員研究費によって行われた成果である。

1) 『財務会計の概念フレームワーク』第1章6項

不足を補うしかないと考えられる。後者であるならば、即応的に解釈情報を入手しなければならない。その際、新聞情報は印刷され配達されるまで待たなければならないばかりか自分が必要とする情報を含んでいるとは限らないし、有料情報は有料たりえるよう情報が作成されるまでの時間を要し決算短信発表直後に入手できるとは限らない。一方、インターネット上の情報は、提供されるサービスの種類に応じた特徴を持つため、条件に適合した情報を入手できる可能性が生じる。

解釈情報や分析能力を十分にもたない個人投資家がインターネット上の情報を利用して不足を補うと想定するためには、そもそも該当する情報をインターネット空間から入手することができるのか、また、入手可能であった場合にどのような情報が共有されるのかを確認する必要がある。

たとえば、SNS（ソーシャル・ネットワーキング・サービス）の中でも Twitter<sup>2)</sup> は、利用者も多く、利用者はフォロー/フォロワーの関係やブロックやミュートなどの機能で繋がる相手をコントロールでき、ニュースの共有や個人の意見や感想の共有がなされやすいサービスである。その反面で当該サービスを利用している必要があるとともに1記事に140文字以内という制約をもつ。会計情報に関しては、Twitter 上では決算短信発表ニュースなどの情報共有がされており、決算情報の入手に貢献していると考えられる<sup>3)</sup>。また、SNSの普及以前から存在するBBS（インターネット掲示板）は、設定されたテーマに関する情報や投稿者の意見や感想を共有することができる仕組みであるが、SNSの普及によってその需要は減少している。BBSは不特定多数の人々が閲覧し投稿できる掲示板であるため、利用者は繋がる相手をコントロールできないが、それ故に相手を特定せずに設定されたテーマについての情報を閲覧することが可能である。また、BBSによっては当該BBSを提供するサービスの利用を開始しなくても情報にアクセスすることが可能であり、Twitterのような文字数制限もないことが多い。現在はSNSほどの利用者数が多くはないが、特定企業の株式取引といったテーマ設定があり文字数の制限なくリアルタイムで投稿できる点では、会計情報に関するより多くの解釈情報が共有される可能性がある。そこで本稿では、決算短信の発表前後の期間にインターネット掲示板でどのような情報が共有され、共有された情報によってどのような影響が生じるのかについて探るための手法、具体的にはテキストマイニングによる調査の可能性と限界を考察する。

## 1. テキストマイニング

### (1) 会計情報に関する疑問と検証手法

財務諸表の役割のひとつとして、投資家による意思決定に資する情報を提供することが挙げ

---

2) Twitter社が提供するSNS

3) 富田 [2018] 参照

られるが、そもそもの疑問として、財務諸表が提供する会計情報は投資家に利用されているのかということが想起される。この疑問については、会計情報が公表されたときに会計情報の内容に応じて株価が変動するか否かを検証することで、間接的に確認することができる。この種の検証は、1960年代以降行われており、多くの実証研究によって利益情報が株価に影響していることは明らかである。どの程度の会計情報が株価に影響するか、あるいはどの程度の会計情報が株価に影響するのかについて、派生研究として現在も展開されている。つまり、間接的な検証ながら、財務諸表が提供する会計情報は投資家に利用されているということが示されている。

次の疑問として、投資家は、財務諸表以外の情報源も利用して企業を評価し投資を行っているため、会計情報は相対的にどの程度の重要性をもって利用されているのかというものがある。投資家にとって会計情報の重要性が低い場合、会計情報の株価への影響が存在しているとしても、コントロールされてようやく検出できる程度の意味しかない。逆に、会計情報の重要性が高い場合、投資家にとっての影響が大きいことを意味するため、会計研究そのものへの重要性も高くなる。富田 [2007] [2008] [2014] では、調査票に基づく結果を利用し、会計情報の参照順序や AHP (Analytic Hierarchy Process, 階層分析法) を使用した会計情報の重要度の推定がなされており、投資家の環境によって依存するところはあるものの、概ね会計情報が他の情報に比して相対的に重要性を持っていることが確認されている。

第三に、ディスクロージャー制度によって会計情報はすべての投資家が利用可能であるが、会計情報の内容を理解するために必要な知識やそのために必要となる情報の収集能力は、投資家によってばらつきがあり、個人投資家の中でも十分な会計情報に対する解釈能力や分析能力をもたない個人投資家はその不足を補う手段を持っているのかという疑問が生じる。不足を補う手段がない場合、その投資行動はギャンブル同様の行動に類似していくため、最終的には証券市場から淘汰されることになると考えられる。逆に、証券市場から淘汰されていかなければ、不足を補う手段が存在するという推測が成り立ちそうである。すなわち、会計情報に対する解釈能力や分析能力が相対的に高い個人投資家から低い投資家への情報移転の存在が想像される。富田 [2015] [2016] [2018] では、社会ネットワーク分析の手法を利用した Twitter での投稿の分析では、利用者の分布、決算短信情報ネットワークの特徴、情報ネットワーク内の会計情報の伝播の存在や株価及び出来高への影響が、部分的に確認されている。

決算短信の発表前後において、決算短信で発表される情報の伝播が部分的に存在しているとしても、どのような情報が伝播し、投資家間でどのような情報が共有されているのかという点については疑問が残る。さらに、共有された情報による投資家の意思決定への影響があるのかという点についても疑問が残る。その確認のためのひとつの手法として、インターネット上での投稿記事に対するテキストマイニングを行い、頻出される用語や共起される用語から共有される情報を推測することが考えられる。また、極性付与（感情分析）を加え、同時期の株価変動と比較することで投資意思決定への影響も推測できるのではないかと考えられる。そのため、

以下ではテキストマイニングという手法について考察してみたい。

## (2) テキストマイニング

「テキストマイニングの研究には大きく分けて二つのアプローチ」<sup>4)</sup>があり、「一つはデータマイニングの研究から派生して、データマイニングの手法を数値データだけでなくテキストデータにも適用されるアプローチ」<sup>5)</sup>であり、「もう一つのアプローチは自然言語処理の研究をベースとしてテキスト分類や情報検索を発展させてテキスト集合から知識発見につなげるアプローチ」<sup>6)</sup>である。前者の「最も単純なとらえ方は、データマイニングの『データ』部分が『テキスト』に置き換わった方法論である、というものであろう。つまり、データマイニングが構造化データから知識発見を行うのに対して、テキストマイニングは非構造化でデータであるテキストから知識発見を行う。非構造化データであるテキストを構造化データに変換するのが自然言語処理の役割で、データが構造化された後は、データマイニングに帰着される、という構図である。」<sup>7)</sup> また、後者のアプローチに従えば、「『大量のテキストデータから自分が必要としている情報を得る』ためには検索技術が存在し、『大量のテキストデータに何が含まれているのかの内容的な分布を知る』ためにはクラスタリングやクラシフィケーションを使った分類整理の技術が存在する。」<sup>8)</sup> これらの技術によって抽出されたテキスト集合から新しい知見を見つけることになる。

しかし、「技術的には、テキストマイニングは情報抽出（自然言語処理）・マイニング（統計処理）・視覚化から構成されることが多く、データマイニングに自然言語処理を組み込んだものという捉え方をされること」<sup>9)</sup>があり、これが広義のテキストマイニングとされている。ところが、「テキストから情報抽出したアイテムを既製のデータマイニングツールで分析しても有用な結果に結びつかないことが多い」<sup>10)</sup>。そのため、狭義のテキストマイニングでは、「テキスト集合から知識発見につなげる」<sup>11)</sup>を強調し、「複数のテキストを分析することで初めて得られる知見を得ることを目的とした技術」<sup>12)</sup>とされる。いずれも言語処理学からのアプローチであるが、「広義のテキストマイニングは、大規模なテキストデータの取り扱いや意図したようなデータの確認を効率的に行うための技術まで含んでおり、新しい知識の発見という意味合いは相対的に弱

---

4) 長谷川 [2001], 893頁

5) 同上, 893頁

6) 同上, 893頁

7) 田中 [2017], 360頁

8) 那須川 [2009], 275頁

9) 同上, 278頁

10) 同上, 278頁

11) 長谷川 [2001], 893頁

12) 那須川 [2009], 275頁

い。一方、狭義のテキストマイニングでは、その目的に知識発見を明確に据えている」<sup>13)</sup>といえよう。

「テキストマイニングは情報抽出、テキスト分類、要約などの自然言語処理と情報検索やデータマイニング、機械学習の組合せで実現される」<sup>14)</sup> わけであるが、ここで、「自然言語処理は、この自然言語の計算機による処理を探索する学術分野」<sup>15)</sup> であり、「自然言語処理の基礎技術は、言語の基本的な情報処理を行うもの」<sup>16)</sup> で、「形態素解析・構文解析・意味解析などがあげられる。」<sup>17)</sup> 「言語を計算機で処理することを探索する自然言語処理は、テキストマイニングに関する言説で、比較的固定的に要素技術を提供する学術分野として述べられることが多い。しかし、実際はテキストの蓄積から分析、抽出した知識断片の解釈まで、かなり柔軟にその技術を変容させ、発展している。」<sup>18)</sup> 「自然言語処理では、文章が与えられ、それらの構成素を認定・分解し、構成素間の修飾・被修飾や意味などの関係を明らかにすることを言語解析とよぶ。自然言語処理で最も盛んに研究されている方向性で、テキストマイニングでも頻繁に利用される形態素解析や構文解析などは、この言語解析に含まれる。」<sup>19)</sup>

最も単純な言語解析は、文単位で、表層である「文」から形態素解析を通じて形態素情報へ、形態素情報は構文解析を通じて構文情報（構文構造）へ、構文情報は意味解析を通じて意味情報（意味構造）へと深層部分に解析を進めていく<sup>20)</sup> ことになる。ここで示される形態素解析について、「言語の意味ある最小の構成素という『語』を想像しそうだが、言語学では言語の『形態素』とよばれる言語単位がそれにあたる。語と形態素は厳密には異なる言語単位である。しかし、自然言語処理では、後処理で形態素を加工するようなこともあり、厳密に区別せず扱うことも多い。」<sup>21)</sup> そして、「文の構文構造を認定するのが構文解析である。構文構造は与えられた文法によって規定される。」<sup>22)</sup> 「語間の関係は、たがいの語義を明瞭にする有力な手がかりで、次の意味解析にも関連する。」<sup>23)</sup> 最後に、「文の意味構造を認定するのが意味解析である。」<sup>24)</sup> 「論理式で記述する方法や、述語を中心とした項構造で記述する枠組みなども与えられているが、形

---

13) 田中 [2017], 352頁

14) 長谷川 [2001], 893頁

15) 田中 [2017], 353頁

16) 同上, 354頁

17) 同上, 354頁

18) 同上, 351頁

19) 同上, 353頁（「一方、発話・発信すべき情報から文や文章を自動産出することを言語生成とよぶ」（田中 [2017], 353頁）

20) 同上, 354-355頁

21) 同上, 355頁

22) 同上, 355頁

23) 同上, 355頁

24) 同上, 355頁

態素情報や構文構造ほど固まっていないのが現状である。」<sup>25)</sup>「近年は、このような構文解析も比較的簡易に適用でき、構文構造を表すデータ構造(木構造)の計数環境も整いつつ」<sup>26)</sup>あり、「テキストマイニングでも積極的に利用されることが予想される」<sup>27)</sup>が、最も単純な言語解析は、「文単位で段階的に進められる」<sup>28)</sup>ため、「文章の解析であっても文ごとに独立に解析されるため、文をこえる文脈のような情報は必然的に失われる。」<sup>29)</sup>また、「各解析が独立に位置付けられているため、ここに研究開発がしやすく、また利用の際も、必要な解析のみ適用したり、同じ言語レベルでプログラムをそっくり入れ替えられるなどの利点がある。他方、言語情報の流れが表層から深層レベルへほぼ一方通行で、各解析の結果が相互に作用しにくい、という問題もある。」<sup>30)</sup>

具体的な解析においては、「語の抽出方法として、Term Weightingによる重要語抽出と、固有表現抽出(Automatic Term Recognition)がある。前者は、統計的に重要である語を抽出するStatistically-orientedアプローチであり、後者は、言語的な知識を中心としたLinguistically-oriented〔アプローチ〕である」<sup>31)</sup>。「テキストを統計的に分析する上では、テキストの文字列全体を一つの値として扱うことは困難であるため、基本的には、テキストに対して自然言語処理を行い、その結果得られるキーワードを抽出して統計処理を行うことになる。」<sup>32)</sup>「大量のテキストデータからの知識獲得の技術はテキストを単に単語の集合として扱う(bag-of-word techniques)ものがほとんどである」<sup>33)</sup>が、「実際のテキストの内容を考慮すると、単純な形態素解析の結果得られる単語の集合のみで、実際の内容を表現することは困難である。」<sup>34)</sup>

「自然言語処理が織り込まれたテキストマイニングの難しさは、マイニングの対象が構造化/非構造化といったデータ形式だけの問題ではない。まず、データマイニングと大きく異なるのは、自然言語処理を通すことによって、人にとって当たり前の重要な言語情報が多く失われることである。」<sup>35)</sup>「文単位の基礎解析を通してしまうと、文脈や状況は必然的に失われてしまう」<sup>36)</sup>のである。そのため、「これらの手法に限ることではないものの、自然言語処理の各種技術をテキストマイニングに導入する場合には、諸技術の基本的な設計方針、こういった言語情報を活

---

25) 同上, 355頁

26) 同上, 355頁

27) 同上, 355頁

28) 同上, 354頁

29) 同上, 354頁

30) 同上, 355頁

31) 長野ほか [2000], 32頁

32) 同上, 32頁

33) 同上, 32頁

34) 同上, 32頁

35) 田中 [2017], 361頁

36) 同上, 361頁

かし、なにを捨て去ってしまい、抽象化してしまっているのか、といったことをきちんと意識することが不可欠である。」<sup>37)</sup> また、「テキストマイニングをより柔軟で巧みな方法論に精密化していくためには、対象領域や課題にそくした自然言語処理の活用と調整、ときには新技術開発も求められる」<sup>38)</sup> のである。

### (3) 会計情報に関連するテキストマイニングの先行研究

前述のとおり、文章の中の文から意味ある語を抽出し、新たな知識発見につなげるものであるテキストマイニングであるが、会計情報に関連するテキストマイニングを利用した研究については、新聞記事や有価証券報告書などを対象として、企業業績あるいは企業の特徴となる意味ある語もしくは文（因果関係文）を抽出する研究と、抽出された語と株価との関連を検証する研究とに分けられる。ここでは、これらの先行研究について確認していきたい。

新聞記事、有価証券報告書、決算短信を対象にテキストマイニングを行い、企業業績あるいは企業の特徴を示す、意味ある語もしくは文（因果関係文）を抽出することを旨とした研究として、竹内ほか [2008]、酒井ほか [2011]、酒井ほか [2015]、佐久間ほか [2018]、佐藤ほか [2018] などが挙げられる。これらの研究は、異なる特徴をもつ2つの企業群からそれぞれに共通する語を抽出するものであったり、業績要因に関連する語（もしくは文）を抽出するものであったりする。その際の工夫として、形態素解析において適合しやすい辞書を選定したり、企業のウェブサイトから業績要因に関する企業固有の情報を集めた辞書を作成したりしているようである。このような形態素解析の工夫と形態素情報を所与とした構文解析によって、意味ある語や因果関係文の抽出に一定程度成功していると思われる。しかし、抽出される語や文の内容は、財務諸表や決算短信で示される数値情報によって示されているものも多く、テキストマイニングによってのみ抽出される内容であるとは言い切れない部分も多い。また、長文で示される内容を同時に多く確認する際の注意すべき語を抽出できているという水準にとどまっているという印象を受ける。今後の進展により、テキストマイニングという方法の選択によって得られる知見が示されるであろうと期待される。

もう一方の方向性として、テキストマイニングによって抽出された語と株価との関連を検証する研究がある。すなわち、言語解析によって抽出された語が新しい知見をもたらすのかという点を、株価との比較で行うテキストマイニングであるといえよう。これらの研究は、テキストマイニングの対象が、ニュース、アナリストレポート、インターネット掲示板、SNSに分類される。以下では、それぞれの対象ごとに分けて先行研究を簡単に確認していきたい。

まず、ニュースと株価との関連に関する研究として、沖本ほか [2014]、阿萬 [2014]、五島

---

37) 同上, 362頁

38) 同上, 362頁

ほか [2016] などが挙げられる。これらの研究では、現在の傾向として、抽出された語の出現頻度との関係を除き、ニュース記事から抽出された語と株価の関係は見いだせないようである。その理由は、ニュースが既に生じた事象を記載するため、株価から見れば結果の後追い情報である、もしくは、ニュースとして報じられた事象はすでに株価に織り込まれているなどのようである。ただし、検証方法をより精緻な方法に発展することで株価との関連性が確認される可能性は、現在のところ否定できないと思われる。

次に、アナリストレポートと株価との関連に関する研究として、高橋ほか [2007]、平松ほか [2018]、鈴木ほか [2020] などが挙げられる。これらの研究において、アナリストレポートによるテキストマイニングで使用する辞書の作成には語の抽出に効果があり、当該辞書を使用することによってアナリストレポートから抽出された語と株価には一定の関連があることが示されている<sup>39)</sup>。しかし、抽出された語は「利益拡大」「業績情報修正」といったものであり、これらのほとんどは数値情報を文字化したものである。すなわち、数値情報（利益情報）としての増益や減益といった情報が既に株価との関連が確認されており、アナリストレポートから抽出された語（文字情報）と株価との追加的な関連性が厳密には明確ではない結果もある。とはいえ、内容の良い（ポジティブ）情報と内容の悪い（ネガティブ）情報に極性を付与した場合、極性が株価リターンに影響しているという結果もあり、極性付与を含め今後の研究に期待されるところである。

第三に、インターネット掲示板における投稿文と株価との関連に関する研究として、丸山ほか [2008]、和田ほか [2012]、柿本ほか [2016] などが挙げられる。これらの研究では、株価の上昇や下落との関係が見いだせていないものの、掲示板への投稿数と出来高やボラティリティとの関係が確認されている。また、相場が変動するときには株式銘柄の議論が活発になり、投稿率が上昇し、一方で相場が安定するときには株式銘柄についての議論は多くなされず、株式銘柄に関係のない投稿の投稿率が上昇する<sup>40)</sup> 傾向があるとされている。これらの結果は、掲示板への投稿と投資家の投資行動との間に関係があることを示唆している。ただ、インターネット掲示板での書き込みは、株式投資に関する用語のほかに、ネットスラングが利用されることもある。利用されるネットスラングは株式投資に関連するものやそれと関係のないものもあるだけでなく、ネット用語としても定着していない表現や文脈によって意味が異なる語も混在している。そのため、インターネット掲示板を対象としてテキストマイニングを行う際に、専門辞書を使うか一般辞書を使うかの選択だけでなく、専用の辞書をどのように構築するかが課題になると考えられる。

最後に、SNSでの投稿と株価との関連に関する研究として、Bollen et al. [2011]、佐藤ほか

---

39) ただし、当該検証では、アナリストレポートで示される意見文と非意見文とを分類する意味はあまりなさそうである。

40) 柿本ほか [2016], 175頁



[2016]、増井ほか [2018] などが挙げられる。Bollen et al. [2011] では、株式市場のデータのみを使った場合よりも、Twitterでの投稿をもとにした感情情報を加えた場合の方が、株式市場（ダウ平均株価）の予測精度が上がるという結果が示されている。また、その他の研究でも、株価分析において、Twitterでの投稿を利用した感情分析は一定の効果があるとされており、SNSが会計情報やその他の情報を補完する、もしくは補助的に機能する可能性を示唆している。ただし、SNSを対象とした研究は、Bollen et al. [2010] の影響を受けているためか、SNSでの投稿記事を基にした感情分析と株価という部分に焦点があたり、その前段階の研究、例えばSNS内での語の内容や抽出された語と株価との関連についての研究がほとんど見られない。

以上から、会計情報に関連するテキストマイニングの先行研究では、テキストマイニングによるより精緻な検証方法を模索しつつ、一定の成果が出始めているようであり、とくに現在のところ、企業の特徴や業績、株価と関連がありそうな語もしくは因果関係文を抽出することに注力しているようである。一方で、テキストマイニングの結果として得られた内容の抽出、数値情報では示されない定性情報を抽出及び当該情報と株価との関連を示すまでには至っていないようである。これは、テキストマイニングに係る研究を待つ必要がある部分もあるが、会計情報に関連する定性情報が、会計情報（数値情報）と重複することも多いことによる難しさにも起因すると考えられる。すなわち、先行研究で示されているテキストマイニングによって抽出された語は、財務諸表による数値情報で示されていることが多く、数値比較で容易に判明する内容となってしまっている。そのため、定性情報に含まれる追加的な情報や意味を抽出する必要が生じている。たとえば、インターネット掲示板で「増益」「減益」という語が抽出され、これらの語が株価と関連があるという結果は、既に会計情報として表現されており、利益情報が株価と関連があることを異なるデータで確認していることになっていると想像される。インターネット掲示板の投稿に「増益」という語が多い増益企業と「増益」という語が少ない増益企業との比較、増益企業であるにもかかわらず「増益」以外の語が多くなる状況及びそれによるインターネット掲示板利用者の関心事の抽出といった付加的な意味をもつ検証が必要であると考えられる。会計研究において、増えてきているとはいえ、テキストマイニングの技法を利用した研究はまだ多くはなく、また同時に、テキストマイニングの技法が自然言語を対象としており、なおかつ日本語を対象とすることから、今後の技法自身の発展と研究の深化を待つ必要がある。

## 2. 決算短信発表前後のインターネット掲示板の投稿に対する試験的分析

### (1) 考察対象

決算短信の発表前後において、決算短信で発表される情報の伝播が部分的に存在しているとしても、どのような情報が伝播し、投資家間でどのような情報が共有されているのかという点

について確認する必要がある、そのためのひとつの方法として、インターネット上での投稿記事に対するテキストマイニングを行い、頻出される用語や共起される用語から共有される情報を推測することが考えられる。インターネット上での情報の伝播も、ニュース記事による伝播、SNSによる伝播、BBSによる伝播などが挙げられるだけでなく、一方向のものも双方向のものも存在する<sup>41)</sup>。富田 [2018] では、SNSのうちTwitterに着目し、決算短信発表前後における会計情報の伝播と共有について考察している。Twitterでの情報伝播は、ひとつの投稿が140文字という制約をもち、Twitterというサービス内でフォローとフォロワーという方向をもった関係の中で行われる。Twitterは利用者数が多く<sup>42)</sup>、利用者に関心のある情報を即時に入手が可能となるが、当該アカウントをフォローしておかなければならない。一方、BBS(インターネット掲示板)はSNSほど利用者数が多いとは言えないかもしれないが、掲示板にテーマ設定があり、文字数の制限なくリアルタイムで投稿できるため、特定の企業の決算短信に関する情報がより多く共有される可能性がある。そこで、インターネット掲示板での決算短信発表前後における会計情報の伝播と共有について考察する必要があるが生じる。

本稿の目的は、その考察のために、インターネット掲示板での投稿を試験的に観察し、テキストマイニングによる調査を行う場合の展望と限界を考察することである。そのために、日本経済新聞社及びYahoo!ファイナンスでの示された時価総額が1位であるトヨタ自動車と2位であるソフトバンクグループを対象<sup>43)</sup>とし、年次の決算短信発表日前後の日において、Yahoo!ファイナンス・株式のそれぞれの企業に設置された掲示板に投稿されたコメントを考察することとした。

## (2) トヨタ自動車の場合

トヨタ自動車の2020年3月期の決算短信の発表は、2020年5月12日13時になされており、発表された内容は、通期ベースでの決算内容はほぼ前年並み(前期売上高30兆2256億円、当期売上高29兆9299億円、前期営業利益2兆4675億円、当期営業利益2兆4428億円、前期税引前当期純利益2兆2854億円、当期税引前2兆5546億円)であるが、次期の業績予想は売上高(24兆円)と営業利益(5,000億円)のみであると同時に約2兆円の減益予想である。なお、次期の予想「税引前利益および親会社の所有者に帰属する当期利益については、新型コロナウイルスの世界的な感染拡大の影響により、現時点では合理的に算定することが困難なため、未定とします」<sup>44)</sup>とされた。

---

41) 富田 [2016] を参照

42) 国内利用者はLINEが最も多く次いでTwitterである。ただし、LINEはSNSというよりコンタクトツールの側面が強いと考えられる。

43) 両方とも2020年8月31日現在の順位はトヨタ、SBG、キーエンス、ソニー、NTTである。

44) トヨタ自動車2020年3月期決算要旨

また、株価については、決算短信発表前日の5月11日は始値6,620円、高値6,675円、安値6,582円、終値6,658円であり、決算短信発表当日の5月12日は始値6,610円、高値6,655円、安値6,485円、終値6,527円であり、決算短信発表翌日5月13日は始値6,412円、高値6,460円、安値6,372円、終値6,384円であり、5月14日は始値6,300円、高値6,326円、安値6,243円、終値6,243円であった。一方、出来高については、5月11日は5,180千株、5月12日は7,809千株、5月13日は8,132千株、5月14日は6,495千株であった。株価は決算短信発表後、一旦上昇したがその後下降しており、出来高は決算短信発表日と翌日にその前後より増加している。

そして、Yahoo! ファイナンス・株式でのトヨタ自動車の掲示板での投稿については、決算短信発表日以前の投稿数はおよそ20～30件であり、決算短信発表日当日の発表後の投稿数は427件、翌日の投稿数は約270件、その後半減している。決算短信発表に関連する投稿の内容は、決算内容に対しては、株価は織り込み済みであり、次期の業績予想に懸念を示すものであった。また、出現数の多かった語<sup>45)</sup>は表1の示すとおりである。株価や株式売買の意思決定に関する語、「株価」「上げ」「下げ」「売り」「買い」が最も多く、次いで決算や利益に関する語、「決算」「利益」「増益」「減益」「赤字」<sup>46)</sup>「予想」が多い。また、「中国」「コロナ」といった時事に関する語が散見されている。一方で、配当に関する語、「配当」「無配」といった語はあまり使用されていない。これらから、決算発表前後では、決算発表そのものや利益情報に関する話題が取り上げられ、その結果を受けた株価の動向、発表された決算内容や株価動向に対応する株式売買の意思決定に関する話題が多いことがわかる。また、決算発表後にそれらの話題が急に増加している<sup>47)</sup>。

表1 決算発表前後における頻出語

	前日	当日発表前	当日発表後	翌日		前日	当日発表前	当日発表後	翌日
投稿数	19	14	427	269	株価に関する語	7	13	160	105
利益に関する語	8	20	135	52	株価	3	0	29	27
決算	6	5	45	17	上げ	1	2	20	5
利益	0	1	37	8	下げ	2	0	28	15
増益	0	0	3	0	売り	0	6	40	24
減益	1	0	15	2	買い	1	5	43	34
赤字	1	0	11	10	時事に関する語	4	0	63	38
予想	0	0	24	15	中国	1	0	18	11
配当に関する語	2	0	16	3	コロナ	3	0	45	27
配当	1	0	10	3					
無配	1	0	6	0					

45) テキストマイニングのためのソフトウェアを使用した結果ではなく、語彙検索で抽出したものである。

46) 「損失」という語は使われていなかった。

47) 投稿された内容の印象としては、ニュース記事の引用を除き、決算内容の詳細や解説を示した投稿は多くはないようであるが、その部分に関する情報は記事の引用によって共有していると思われる。

### (3) ソフトバンクグループの場合

ソフトバンクグループの2020年3月期の決算短信の発表は、2020年5月18日15時<sup>48)</sup>になされており、発表された決算内容は、通期ベースでの売上高はほぼ前年並み（前期売上高6兆935億円、当期売上高6兆1851億円）であり、利益は大幅な減少（前期営業利益2兆736億円、当期営業損失1兆3646億円、前期税引前当期純利益1兆6826億円、当期税引前355億円）である。また、次期の業績予想は開示されていない。営業利益の減少は、「ソフトバンク・ビジョン・ファンド等SBIAの運営するファンドからの営業利益」が前期は1兆2566億円であったが、当期は△1兆9313億円であったためである。ただし、この影響は、第1四半期に計上された「アリババ株式先渡売買契約決済益（1.2兆円）」によって軽減されている。

また、株価については、決算短信発表前日（営業日ベース）の5月15日は始値4,600円、高値4,615円、安値4,501円、終値4,574円であり、決算短信発表当日の5月18日は始値4,614円、高値4,713円、安値4,572円、終値4,621円であり、決算短信発表翌日の5月19日は始値4,697円、高値4,784円、安値4,425円、終値4,494円であり、5月20日は始値4,445円、高値4,589円、安値4,384円、終値4,539円であった。一方、出来高については、5月15日は16,681千株、5月18日は25,380千株、5月19日は41,866千株、5月20日は27,771千株であった。この期間の株価は4,600円付近で始まり、一日の間で上げ下げした後、4,500円付近で終了する傾向であり、出来高は決算短信発表日翌日のみ多くなっている。

そして、Yahoo! ファイナンス・株式でのソフトバンクグループの掲示板での投稿については、決算短信発表日前日の投稿数は約500件弱であり、決算短信発表日当日の発表前の投稿数は約1,000件、決算短信発表日当日の発表後の投稿数は約2,200件、翌日の投稿数は2,430件であり、トヨタ自動車と比べて多くの投稿がある。決算短信発表前後に投稿で示された特徴的な話題は、次のようなものである。決算短信発表前日（17日）11時過ぎに、サウジ PIF（パブリック・インベストメント・ファンド）がビジョン・ファンドへの投資を担保に借入を行うというニュースが共有されており、またソフトバンクグループが1兆円の赤字があり得るという投稿者の予測も共有されている。決算短信発表日（18日）の発表前8時40分ごろに、アリババに関連するニュースやジャック・マー退任のニュースが共有され、13時30分ごろ T モバイル売却協議がなされていることのニュースが共有されている。また、最大2兆5000億円のうち5000億円の自社株買いを実施すること及び決算予測修正のニュースが共有されている。決算短信発表当日の発表後は、まず第4四半期で営業損失がおおよそ1.3兆円に拡大したことが話題になり、その所感が示された。次に、インターネットでライブ中継された決算説明への感想が多く示されており、孫氏の表情から読み取った決算内容との関係と感想、ユニコーン発言、ビジョン・ファンド保有企業88社のうち15社が倒産、ウィーワークへの反省発言などへの揶揄を含めた感想お

---

48) 16時30分より記者発表があり、インターネットでライブ配信されている。

よびソフトバンクグループの倒産の可能性の有無、無配の可能性、アリババ株式売却デリバティブなど議論や感想などが投稿されている。そして、19時ごろ以降に改めて決算内容の感想の共有や自社株買いの是非に関する主張などが共有されている。決算短信発表翌日は、決算の内容と自社株買いへの解説や感想の共有、それらを受けた株価変動やデイトレーディングの結果や対応の投稿がなされている。

また、出現数の多かった語<sup>49)</sup>は表2の示すとおりである。株価や株式売買の意思決定に関する語、「株価」「上げ」「下げ」「売り」「買い」が最も多く、決算発表や決算説明の影響を受けた株価変動やそれに伴う株式売買に関する書簡や情報共有が多くなされていることがわかる。また、出現数は多くはないが、決算発表当日の発表前や決算発表翌日に「利確」が散見され、決算発表の影響を避けるあるいは決算発表の影響後すぐに売却し損失を回避しようとしている場合があり、その状況を共有していることが想像される。さらに株価への「(織り)込み」を示す語も散見されており、決算の内容が予測され、ある程度株価に反映されているか株価動向を予想していることを示唆している。次いで決算や利益に関する語、「決算」「利益」「増益」「減益」「赤字」「予想」が多い。決算発表当日には「決算」が多く、決算発表後には「赤字」が多い。決算発表当日は決算内容への関心が高まっており、その後は発表された内容に関する感想や情報共有が多くなされていることがわかる。ここで、「損失」より「赤字」と表現されていることが多く、インターネット掲示板においてはより一般的な言葉で投稿されていると考えられる。第三に、時事に関する語「中国」「コロナ」や、ソフトバンクグループに関する語「(ジャック・)マー」「アリババ」などが多い。決算や次期の業績への影響が予想される事象や当該事象への所感を投稿し、情報共有しているようである。また、決算発表当日の発表後に「配当」が多く、決算説明で言及された配当についての感想や反応が投稿されている。

ソフトバンクグループの掲示板では、玉石混交の状態ではあるが、ソフトバンクグループの決算を含む詳細な分析や関連ニュースの共有や感想が多く、トヨタ自動車のそれと比べると平均的に長文で情報量が多い印象である。また、株価変動に敏感であり、関連する投稿も多いようである。

以上から、Yahoo!ファイナンスのトヨタ自動車やソフトバンクグループの掲示板で、該当企業に関するニュース、決算短信で発表された利益情報を中心とする決算内容やそれによる株価動向への所感などが、情報として共有されていることがわかる。このような情報の共有は、他の企業であっても、同様に行われているのか、今後検証していく必要がある。また、掲示板に投稿される投稿数や情報量に違いがある場合、投資家の意思決定にどのような影響が生じているのかについても、今後検証していく必要がある。

---

49) テキストマイニングのためのソフトウェアを使用した結果ではなく、語彙検索で抽出したものである。

表2 決算発表前後における頻出語

	前日	当日発表前	当日発表後	翌日		前日	当日発表前	当日発表後	翌日
投稿数	481	1,054	2,192	2,430	企業に関する語	106	164	295	306
利益に関する語	148	213	537	358	株主価値	14	3	12	48
決算	82	129	195	133	マー	8	74	71	34
利益	9	9	46	29	アリババ	36	56	136	160
増益	1	0	0	0	ビジョン	31	12	47	28
減益	3	0	3	4	スプリント	14	8	12	16
赤字	26	59	216	131	Tモバイル	3	11	17	20
損失	14	8	24	24	時事に関する語	92	54	175	121
予想	13	8	53	37	中国	34	17	43	21
配当に関する語	4	13	187	64	コロナ	58	37	132	100
配当	4	11	150	45					
無配	0	2	37	19					
株価に関する語	283	686	1,009	1,344					
株価	49	91	169	197					
上げ	22	75	114	127					
下げ	22	28	80	105					
売り	46	155	242	397					
買い	131	320	338	489					
(織り)込み	13	7	35	20					
利確	0	10	4	9					

### 3. テキストマイニングからのアプローチと着地点

本稿は、決算短信で発表された会計情報や会計情報の解釈情報が、インターネット掲示板で共有されているのか、また、具体的にどのような情報が共有されているのか、について検証するための手法としてのテキストマイニングの考察を目的としている。そこで、テキストマイニングについての考察と、決算短信発表前後でのトヨタ自動車とソフトバンクグループのインターネット掲示板への投稿を概観した。その結果、現状においていくつかのハードルが存在していると考えられる。その最大のハードルは用語の問題であろう。

テキストマイニングを実施する際の形態素解析において、辞書の調整が大きく影響する。インターネット掲示板では、多くのスラングや株式投資に関する専門用語が見られる。たとえば、ノンホル、マイ転、ナイアガラ、S安、お父さん、寄り天、地合い、売り豚、PTS、VIX、ABRなどである。これらのうち株式投資に関する専門用語については、対応する辞書を利用することで対処できると考えられるが、スラングについては、特定の文脈において意味を持ち、単純な語句変換では対処できない語があり、語を抽出するだけでは意味がない場合が存在する。ま

た、誹謗中傷に利用される語であったとしても、全体的な傾向を理解するために重要であったりするため、単純に除外すればよいというものでもなさそうである。単純にスラングや誹謗中傷とされる語を除外してしまうと、「上位に抽出される特徴的な表現は当たり前のものであることが多い」<sup>50)</sup> くなり、「このような場合、既知の事実ばかり得られ、今まで気がつかなかった有用知見が見つかることは少ない」<sup>51)</sup> ことになる。また、会計情報の内容を示す言葉が一般に利用される語で示されること方が多いこともある。たとえば損失とは表記せず、赤字と表記されることが多い。そのため、決算短信発表前後における会計情報の共有について、インターネット掲示板を対象としてテキストマイニングを行う場合、株式投資に関する専門用語は辞書に反映させるとしても、その他については、一般的な語の辞書にスラング部分を追加した方が抽出精度は上がりそうである。ただし、企業固有の事情によって出現する語、すなわち、企業固有の情報については、企業ごとに辞書を作成するか、一旦、企業ごとに形態素解析を行い企業固有の事象を抽出したのちに、インターネット掲示板全体で抽出するかといった工夫が必要となる。さらに、「分析に有効な観点の設定や辞書の準備が分析の結果に影響する」<sup>52)</sup> ため、「有用な分析をするためには分析者の経験や勘が必要となる課題」<sup>53)</sup> があり、インターネット掲示板での投稿内容や傾向をある程度理解しておく必要がある。

これらの対処によって、テキストマイニングを使用した語の抽出とその頻度や共起性を時系列に従って検証するによって、共有される情報を推測できると考えられる。また、企業固有の事象について共有される情報は、企業ごとに抽出された語の頻度を検証することで推測できると考えられる。さらなる検証として、共有された情報もしくは抽出された語の内容が、投資行動に影響したか否かについては、結果として株価形成や出来高に影響すると考えられるため、株価や出来高との関連を検証することで推測できると考えられる。その際、先行研究の結果を踏まえると、単にテキストマイニングによって抽出された語と株価や出来高との関連を検証するのではなく、会計情報で示される数値情報を組み合わせた上での株価や出来高との関連を検証する必要がある。ただし、抽出された語による感情分析や極性の付与を行い、その結果と株価や出来高との関連を検証するには、テキストマイニングそのものの研究が進展するまで待つ必要があると考えられる。

#### 参考文献

- 阿萬弘行 [2014], 「公開情報・マスメディア報道が株式市場へ及ぼす効果についての考察」, 『商学論究』, 62巻 1号, 61-79頁
- Bollen et al. [2010], "Twitter mood predicts the stock market," *Journal of Computational Science*, Vol.2,

---

50) 竹内ほか [2008], 1頁

51) 同上, 1頁

52) 同上, 1頁

53) 同上, 1頁

No.1, pp.1-8

- 五島圭一、高橋大志 [2016], 「ニュースと株価に関する実証分析——ディープラーニングによるニュース記事の評判分析——」, 『証券アナリストジャーナル』, 54巻3号, 76-86頁
- 長谷川隆明 [2001], 「テキストマイニング」, 『人工知能学会誌』, 16巻6号, 893-896頁
- 平松賢士、酒井浩之、坂地泰紀 [2018], 「深層学習を用いて極性付与されたアナリストレポートと株式リターンとの関連性」, 『第20回人工知能学会金融情報学研究会』(資料)
- 柿本研人、諏訪博彦、小川祐樹、梅原英一、安本慶一 [2016], 「インターネット株式掲示板における話題と株式指標の関係」, 『マルチメディア、分散協調とモバイルシンポジウム2016論文集』, 170-176頁
- 北森詩織、酒井浩之、坂地泰紀 [2016], 「決算短信PDFから業績予測文の抽出」, 『言語処理学会 第22回年次大会発表論文集』, 663-666頁
- 小寺俊哉、佐藤史仁 [2018], 「テキストマイニングによる有価証券報告書の因果関係文以外の特徴文の抽出」, 『日興リサーチレビュー』, 2018年1月号, 1-15頁
- 小寺俊哉、佐藤史仁、佐久間洋明、田中良典 [2019], 「テキストマイニングを用いた有価証券報告書からの未来志向文の抽出」, 『日興リサーチレビュー』, 2019年3月号, 1-18頁
- 丸山健、梅原英一、諏訪博彦、太田敏澄 [2009], 「インターネット株式掲示板の投稿内容と株式市場の関連性」, 『第2回人工知能学会ファイナンスにおける人工知能応用研究会』(資料)
- 増井佑亮、藤野巖 [2018], 「Twitterの多軸的感情情報を利用した株価の予測」, 『言語処理学会第24回年次大会発表論文集』, 69-72頁
- 長野徹、武田浩一、那須川哲哉 [2000], 「テキストマイニングのための情報抽出」, 『情報学基礎』(情報処理学会研究報告60回), 31-38頁
- 中山祐輝、津々見誠、村上浩司 [2019], 「文間の結束性に基づく決算短信における業績要因文の抽出」, 『言語処理学会 第25回年次大会発表論文集』, 1351-1354頁
- 那須川哲哉 [2009], 「テキストマイニングの普及に向けて——研究を実用化につなぐ課題への取組み——」, 『人工知能学会誌』, 24巻2号, 275-282頁
- 西野嘉之 [2017], 「有価証券報告書の類似度による企業評価」, 『情報システム学会全国大会論文集13回』(資料)
- 沖本竜義、平澤英司 [2011], 「ニュース指標による株式市場の予測可能性」, 『証券アナリストジャーナル』, 52巻4号, 67-75頁
- 酒井浩之、増山繁 [2011], 「企業業績発表記事を対象としたテキストマイニング」, 『第54回自動制御連合講演会講演論文集』, 466-470頁
- 酒井浩之、西沢裕子、松並祥吾、坂地泰紀 [2015], 「企業の決算短信PDFからの業績要因の抽出」, 『人工知能学会論文集』, 30巻第1号, 172-182頁
- 酒井浩之、松下和暉、北島良三 [2019], 「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」, 『知能と情報』, Vol.31, No.2, 653-661頁
- 佐久間洋明、田中良典 [2018], 「有価証券報告書に含まれるテキスト情報と企業業績の関係について」, 『日興リサーチレビュー』, 2018年3月号, 1-13頁
- 佐藤大吾、大原剛三、豊田哲也 [2016], 「経済情報に関心をもつSNSユーザの投稿内容に基づく株価暴落予測モデルの構築」, 『人工知能学会第110回研究会』(資料)
- 佐藤史仁、佐久間洋明、小寺俊哉、田中良典、坂地泰紀、和泉潔 [2018], 「テキストマイニングによる有価証券報告書からの因果関係文の抽出」, 『第20回人工知能学会金融情報学研究会』(資料)
- 佐藤謙太、小高知宏、黒岩文介、白井治彦 [2015], 「ネガポジ解析によるWebデータと株価変動の相関関係評価」, 『福井大学大学院工学研究科研究報告』, 62・63合併巻, 75-86頁
- 白田佳子、竹内広宣、荻野紫穂、渡辺日出雄 [2009], 「テキストマイニング技術を用いた企業評価分析——倒産企業の実証分析——」, 『経営分析研究』, 第25号, 40-47頁
- 鈴木雅弘、堅木聖也、坂地泰紀、和泉潔、石川康 [2020], 「テキストマイニングによるアナリストレポートを



- 用いた株価動向予測], 『言語処理学会第26回年次大会発表論文集』, 307-310頁
- 高橋悟、高橋大志、津田和彦、高橋正和 [2007], 「テキストマイニングを用いた株式市場におけるアナリストレポート効果に関する分析」, 『経営情報学会全国研究発表大会要旨集』(2007年度春季全国研究発表大会資料)
- 竹内広宣、荻野紫穂、渡辺日出雄、白田佳子 [2008], 「文脈情報を利用した文書集合間の差異分析方法と決算報告書分析への応用」, 『人工知能学会全国大会論文集22回』, 1-4頁
- 田中省作 [2017], 「自然言語処理からみた『テキスト』と『テキストマイニング』」, 『社会学評論』, 68巻3号, 351-367頁
- 和田英一、諏訪博彦、小川祐樹、太田敏澄 [2012], 「専門辞書を用いたテキストマイニングによるインターネット株式掲示板の投稿分析に関する研究」, 『経営情報学会全国研究発表大会要旨集』(2012年度秋季全国研究発表大会資料), 115-118頁
- 山本零、川代尚哉、栗田昌孝 [2020], 「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」, 『ジャファイ・ジャーナル』, 18巻, 46-62頁
- 保田明夫 [2005], 「テキスト・マイニングの概要」, 『電気学会論文誌C(電子・情報・システム部門誌)』, 125巻, 5号, 682-689頁
- トヨタ自動車2020年3月期決算要旨
- ソフトバンクグループ2020年3月期決算短信
- 富田知嗣 [2007], "Investors' Attitude to Information on Equity Investment," *Journal of Accountancy, Economics and Law*, No.1, pp.23-40
- [2008], 「個人投資家における会計情報の役割」『会計』, 第174巻第2号, 54-65頁
- [2014], "The Effect of Stock-Investment Style on the Selection of information," *Journal of Accountancy, Economics and Law*, No.8, 2014年3月, pp.19-39
- [2015], 「株式投資意思決定におけるソーシャル・ネットワークの影響」, 『現代社会と会計』, 第9号, 2015年3月, pp.133-144.
- [2016], 富田知嗣, 「決算発表時のTwitter拡散による証券市場への影響」, 『現代社会と会計』, 第10号, 2016年3月, pp.59-80
- [2018], 富田知嗣, 「決算情報に関するTwitterの利用状況」, 『現代社会と会計』, 第12号, 2018年3月, pp.33-53