

〈研究ノート〉

## 企業の掲示板の投稿と会計情報

### Posts on Company Message Boards and Accounting Information

中村 彰良

#### I はじめに

会計情報と株価に関する研究では、会計情報のグッドニュース、バッドニュースは株価に反映されるものの、その会計情報が公表される以前にかなり株価に織り込まれてしまう傾向が認められる。株価に織り込まれてしまうということは、投資家等が将来公表される会計情報の内容を何らかの方法で察知していると考えられる。一方、企業の掲示板の投稿内容には、その時の投稿者のセンチメント（ポジティブ、ネガティブ）が反映されることが考えられるので、グッドニュースが察知された場合には、ポジティブな投稿が増え、バッドニュースが察知された場合には、ネガティブな投稿が増えるのではないかと考えた。そしてセンチメント分析の手法を用いて、この問題を検討することにした。筆者はセンチメント分析を今までにやったことがないので、とにかくやってみることを重視している。

次のIIにおいて、この研究の背景となる先行研究について述べる。IIIにおいて、研究方法と使用するデータ等について説明する。IVにおいて、分析結果を示す。Vにおいて、今後の課題等を検討する。

#### II 背景

決算情報の情報効果を分析する研究は数多くなされてきている。多くの研究は、会計情報が株価に与える影響をとらえようとするものである。Ball and Brown<sup>1</sup>は、株価の変動による実際の収益率と期待される収益の理論値との差を累積させたabnormal performance index（異常業績指標、以下API）という指標を用いて、会計情報でグッドニュース（増益）を報告することになる企業グループとバッドニュース（減益）を報告することになる企業グループとに分けてAPIの推移をニューヨーク証券取引所上場企業のデータを用いて分析している。結果として、グッドニュースの企業グループのAPIは会計情報を公表する1年前からだいたい連続して上昇していて、バッドニュースの企業グループのAPIは会計情報を公表する1年前からだいたい連続して下降していた。このため1年前に会計情報の内容を知っていれば、超過収益を得ることが可能であ

1 Ball R. and P. Brown (1968) An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, Vol.6, No.2, pp.159-178.

る。ただこのグッドニュースとバッドニュースは会計情報が公表されるまでにかかなり織り込み済みになることもわかった。市場では会計情報が公表される前に、その内容を察知していることが考えられる。

石塚<sup>2</sup>では、Ball and Brownの手法をベースにしなが、日本で公表される経営者による利益の予測情報に注目した研究を東京証券取引所上場企業のデータを用いて行っている。この利益予測情報は、実績の決算情報と同時に公表される。このため株価（したがってAPI）に与える影響が、実績の決算情報によるものなのか予測情報によるものなのか、さらにポートフォリオを分けることで分析を行っている。実績が増益（グッドニュース）か減益（バッドニュース）かだけでグループ分けするのではなく、利益予測が今期より増益（グッドニュース）か減益（バッドニュース）かも組み合わせている。実績が増益で予想も増益の場合GG、実績が増益で予想は減益の場合GB、実績が減益で予想は増益の場合BG、実績は減益で予想も減益の場合BBという4つのグループに分けてAPIの推移を観察している。結果として、GGとGBの企業グループについては、情報が公表される1年前からしばらくはAPIが同じように上昇していくが、情報が公表される3か月ぐらい前からGBの企業グループのAPIは下方に離れていくことが観察された。このことから利益の予測情報に株価が反応していると考えられる。そして利益の予測情報は、それが公表される前にある程度織り込まれることもわかる。一方、BGとBBの企業グループについては、情報が公表される1年前からほとんど同じように下降していることが観察された。この結果から、実績が増益で予測が減益の場合、予測情報が察知されている可能性が高い。ただこの研究では中間決算を考慮していないので、中間決算が芳しくなかった企業が次期の減益予測をする傾向があるとすると、中間決算の内容に反応しているだけかもしれない。

このように次期の利益あるいは利益予測の情報が市場では察知されているようであるが、どうしてそのようなことが可能になるのかについて明確に説明した研究はまだないと思われる。ただ、オルタナティブな情報（例えば、その企業の荷動き）あるいは何らかの事情で、会計情報が公表される前に、その情報が市場参加者に察知されている可能性は十分考えられることである。

一方、近年において情報処理技術の発展にともなって、テキスト分析の手法を用いた研究が、様々な分野に広がっている。会計の分野でも、企業の開示した情報をテキスト分析した研究が増えている。Loughran and McDonald<sup>3</sup>は、こういった研究をサーベイしている。彼らによれば、情報利用者が情報提供者の意図した内容を理解しやすいかという可読性に焦点を当てるものもある。こういった研究では、利益の少ない企業のアニュアルレポートは可読性が低いかどうかといった問題を考察している。他の多くの研究は、テキストから何らかの意味合いを引き出そうとすることに焦点を当てている。こうした研究には、センチメント分析やトピックモデリングや文書の類似性の測定などがある。トピックモデリングでは、発生主義に関連する文章を抽出し

2 石塚博司編著（1987）『実証会計情報と株価』同文館。

3 Loughran T. and B. McDonald (2016) Textual Analysis in Accounting and Finance : A Survey. *Journal of Accounting Research*, Vol.54, No.4, pp.1187-1230.

たり、類似性の測定では、年度による記述内容が変わっているかなどが検討されたりする。しかし会計分野へのこういった手法の適用については、まだ模索中というところもあるようで、多くの研究は、センチメント分析に係るものが多いようである。

会計分野でのテキスト分析は、企業が公表する資料に含まれるテキストを分析対象とするものが多い。しかし投資家等が投資の判断をするときに用いる情報は、企業が公表するものに限定されるわけではないと考えられる。ソーシャルネットワークサービス上で得られる情報やムード（センチメント）も投資家等の判断に影響を与えるものと思われる。ツイッターのツイートに含まれるセンチメントで株価を予見できるかどうかを検討した研究として、Bollen, Mao and Zeng<sup>4</sup>がある。彼らは、ツイートのうち”I feel”といったような言葉が入っている感情を表現していると考えられるものを取り上げている。OpinionFinder (OF)を利用して得たセンチメントとGoogle-Profile of Mood States (GPOMS)を利用して得た6次元 (Calm, Alert, Sure, Vital, Kind, Happy) のセンチメントとDow Jones Industrial Average (DJIA) との関係进行分析している。2008年2月28日から11月3日のデータを使ってグレンジャー因果性検定を行っている。結果として、2日から6日のラグのあるCalmのデータについてDJIAと有意なグレンジャー因果性が認められた。つまり2日から6日前のCalmのセンチメントがわかっているならば、株価の予測に役立つと考えられるということになる。

上記研究は、世間一般のセンチメントを考えているが、特定の会社に係るセンチメントもあると考えられる。特定の会社に係るセンチメントと株価の関係について分析した研究として、Das and Chen<sup>5</sup>がある。彼らは、会社のmessage boardsに投稿されたメッセージのタイプ（センチメント）を「買い」と「売り」と「中立」に分類するために5つの方法を使い、3つ以上の方法で同じ分類になったときにその分類として、3つ以上の方法で同じ分類になることがないメッセージがあったときには、それを捨てている。分類するための5つの方法は、事前に分類されたテストデータでチューニングされている。そして投票型の分類方法でメッセージ（データ）を分類している。データは、Morgan Stanley High-Tech Index (MSH) を構成する24社のmessage boardsに2001年の7月と8月の2か月間に投稿されたメッセージと同じ期間の各社の株価とMSHの値である。ハイテク企業については、投稿が活発に行われているので、このようにしたようである。各社のメッセージは「買い」の場合プラス1とし、「売り」の場合マイナス1として日次で集計している。また全社分を集計したものを統合センチメント指標として計算している。各社のセンチメント値と統合センチメント指標および各社の株価、MSHはそれぞれ標準化されている（時系列の平均値を引き、時系列の標準偏差で割っている）。結果として、統合センチメント指標とMSHは同じような動きをすることが観察された。また、MSHを被説明変数として、1期前のMSHと統合センチメント指標を説明変数とする重回帰分析で、統合センチメント指標の係数の値は0.154で有

4 Bollen J., H. Mao and X. J. Zeng (2011) Twitter mood predicts the stock market. *Journal of Computational Science*, pp.1-8. (<https://arxiv.org/abs/1010.3003>).

5 Das S. R. and M. Y. Chen (2007) Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, Vol.53, No.9, pp.1375-1388.

意なものであった。一方個別の各企業のセンチメント値と各企業の株価の関係については、標準化したデータであることもあり、全社分をまとめて分析を行っている。そして適合度検定では一応有意であったようであるが、回帰分析したときの $R^2$ は0.0041ということで、ほとんど説明力を持っていないことも明らかになった。このように、統合した指標どうしの関係は強いと考えられるが、各社ごとのセンチメントと株価の関係はあまり強くないかもしれない。Das and Chenは、センチメントをクロスセクションで統合することによって統合センチメント指標のセンチメントをとらえる質が改善されていると考えているようである。

このように財務情報と株価の研究では、グッドニュース、バッドニュースが市場参加者に察知されている可能性が指摘され、ネット上の投稿のセンチメントについての研究では、センチメントと株価の関係は認められるようである。通常グッドニュースは株価を上昇させると考えられ、バッドニュースは株価を下落させると考えられるので、ある企業のグッドニュースを察知した者がいた場合、その企業についての投稿のセンチメントはポジティブに寄っている可能性があると考えられる。逆にある企業のバッドニュースを察知したものがいた場合、その企業についての投稿のセンチメントはネガティブに寄っている可能性があると考えられる。YAHOOファイナンスの企業ごとの掲示板の投稿を利用して、こういった傾向があるのか確かめてみることを考えた。

### Ⅲ 分析方法とデータ

ここで問題にするのは、グッドニュース、バッドニュースが公表される前の段階で、企業の掲示板の投稿のセンチメントにそれが反映されているかどうかである。前に取り上げた財務情報と株価に関する研究では、増益をグッドニュース、減益をバッドニュースとしていた。しかし多くの日本企業の場合、利益の予測情報を公表しているので、最終的な利益が利益予想の数値よりも大きかったケースをグッドニュースと考え、最終的な利益が利益予想の数値よりも小さかったケースをバッドニュースと考えることにした。ここで問題になるのが、利益予想の数値は期間中に修正される場合があることである。利益予想の数値に大幅な修正がなされる場合には、その都度適時に開示されることもあるが、四半期の開示の際に変更される場合もある。したがって、第3四半期の開示から本決算の開示までの期間は、利益予想の数値に大きな変化がなければ、企業から利益に係る公式の情報は公表されることはない。この期間に、グッドニュースになる企業群とバッドニュースになる企業群とで掲示板の投稿のセンチメントに違いがあれば、ニュースを察知した者がそれを投稿に反映させた可能性があると考えられる。また察知した情報の確度は本決算が近づくにつれて増すだろうと考えられるので、グッドニュースになる企業群とバッドニュースになる企業群とで掲示板の投稿のセンチメントの時間的推移もどうなるか見てみることにした。

第3四半期の開示から本決算の開示までの期間について企業の掲示板の投稿データを収集することにしたが、この期間は企業によって違いがある。3月決算の企業の場合、

第3四半期の決算短信が公表されるのは1月下旬から2月上旬にかけてのケースが多く、本決算の決算短信が公表されるのは4月下旬から5月中旬にかけてのケースが多い。このため、いつからいつまでの情報を収集しなければならないのかを事前に決定することはできなかった。

前述のように、YAHOOファイナンスの企業ごとの掲示板の投稿を利用することにしたので、2022年1月からこの掲示板の下調べをした。上場企業すべての掲示板があるが、企業によって投稿数にはかなりの差があった。中には何か月に1件程度の投稿しかない企業もあった。当然こういった企業のデータは使えないことになる。そこでどういった企業を選定するかということになる。YAHOOファイナンスには各種のランキングのデータが入手できるようになっているので、ランキングで時価総額上位の企業を見てみると、有名な大企業ばかりであり、その掲示板を見てみると、投稿はかなり活発に行われていた。このため、時価総額上位の企業の掲示板の投稿データを収集することにした。また企業数については、Das and Chenの研究でハイテク企業24社の投稿データを使っていたので、根拠は曖昧であるが、20社の投稿データを収集することにした。基本的には、2022年1月11日の時価総額上位20社を選定したが、20社の中のファーストリテイリングと中外製薬は3月決算ではないので外すことにした。また、20社の中にソフトバンクグループとソフトバンクという親子会社が入っていたので、親会社のソフトバンクグループを入れて、子会社のソフトバンクを外すことにした。20社を維持するため、時価総額のランキングで23位までの企業のうち外した3社以外の投稿データを収集することにした。20社の具体的な企業名は、次節の表に書いてある。

つぎに投稿データの収集頻度について、どうするか検討した。Das and Chenの研究のように、投稿のセンチメントと株価の関係を検討している場合、株価は毎日変わるので、投稿も日次でデータを収集している。本研究では、将来明らかになるグッドニュース、バッドニュースと投稿のセンチメントとの関係を検討しようとするものである。毎日データを収集しなくてもよいだらうと考えた。ただ世間全般のセンチメントに影響のある出来事がある日もあるかもしれないので、20社分の投稿データは、なるべく同じ日の同じ時間に収集する必要がある。またセンチメントの時間的推移も見てみたいので、第3四半期の決算短信の公表が始まる1月末から定期的に投稿データを収集することにした。具体的には、2022年1月24日から2週間ごとに投稿データを収集することにした。20社のうち第3四半期の決算短信の公表が一番遅かったのがリクルートの2月14日で、本決算の決算短信の公表が一番早かったのが日本電産の4月21日であった。このため、その間に挟まれている2月21日、3月7日、22日、4月4日、18日の5日間に収集した投稿データを使うことにした。またこの期間の20社の適時開示の状況についてTDnetで確認してみたが、自己株式の取得状況に関する開示や役員の異動に関する開示はいくつかあったが、利益予想の変更に関するものはなかった。

YAHOOファイナンスの企業ごとの掲示板からの投稿データの収集は、Pythonを使ってスクレイピングすることにした。YAHOOファイナンスの掲示板は、最初のページを最後までスクロールすると次のページに移っていくタイプ（無限スクロール）のウェブサイトだった。このようなタイプのウェブサイトをスクレイピングする

場合、通常はSeleniumというツールを使うのが一般的なようである。しかしSeleniumを利用する場合、Windowsパソコンでは仮想マシンの環境が必要になるようで、もう少し手軽なものとしてRequest-HTMLというPythonのライブラリを利用することにした<sup>6</sup>。何ページ分の投稿を収集するかについては、事前の下調べの段階で、2ページまでであれば、各社おおよそその日に投稿されたものが得られていたことと（投稿は新しいものから表示される）、さらにページを増やそうとするとパソコンの動作が不安定になることもあり、各社2ページずつ投稿を収集することにした。2ページで、投稿は各社だいたい50件程度であり、5日分をまとめると250件程度になる<sup>7</sup>。

収集した投稿のテキストデータを分析する際に、日本語の場合、英語と違って単語の間に空白が入っていないので、単語に分割する必要がある。本研究では、単語に分割するためのツールとしてJanome Tokenizerを利用することにした。

投稿データについてセンチメント分析を行うためには、ポジティブな単語、ネガティブな単語をリスト化した極性辞書が必要になる。奥村<sup>8</sup>によると、日本における有価証券報告書のテキストを対象にしたセンチメント分析（トーン分析）では、Loughran and McDonald<sup>9</sup>の作成したリストを翻訳したものが用いられることが多いようである。今回の収集した投稿には、会計の専門用語がそれほど出てくるわけではないので、より一般的な極性辞書があれば、そちらの方がよいのではないかと考えた。そして本研究では、高村<sup>10</sup>の単語感情極性対応表を用いることにした。この単語感情極性対応表は、高村・乾・奥村<sup>11</sup>によると、各単語を同極性リンク（類義語など）と逆極性リンク（反義語など）で結び付けたネットワークを構築し、少数の極性の明らかな単語から他の単語の極性値を割り当てていく手法によって作成されているようである。この単語感情極性対応表の各単語に与えられる極性値は、マイナス1からプラス1の範囲の値（連続値型）をとる。また、この単語感情極性対応表の極性値は、マイナスが多いことが指摘されている<sup>12</sup>。

#### IV 分析結果

各社の投稿データについて、今回は、活用がなく単独で考えやすいということもあり、名詞だけを取り出し、単語感情極性対応表のその単語の極性値を集計していった（単語感情極性対応表にない単語については、ゼロとしてカウントしていない）。各社の投稿日ごとの集計値とその合計は、表1にまとめられている。マイナスの極性値の

6 WindowsパソコンでSeleniumを利用する場合、通常、仮想マシンのUbuntuを使うことになるが、Request-HTMLの場合、それがPypeteerを使うかたちで、Pythonだけで利用できる。

7 それぞれの投稿の長さは様々であるが、その発生についてはランダムなものと仮定している。

8 奥村雅史委員長（2022）『情報関連技術の進展と財務会計に関する研究』日本会計研究学会特別委員会最終報告書。

9 Loughran T. and B. McDonald (2011) When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, Vol.66, No.1, pp.35-65.

10 高村大也『単語感情極性対応表』([http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html)).

11 高村大也、乾孝司、奥村学（2006）「スピンモデルによる単語の感情極性抽出」『情報処理学会論文誌』Vol.47, No.2, pp.627-637.

12 山内長承（2017）『Pythonによるテキストマイニング入門』オーム社。

単語が多いことから、集計値はすべてマイナスになっている。相対的な値の方がわかりやすいと思い、各投稿日ごとにデータを標準化することにした（平均値を引き、標準偏差で割った）。標準化されたデータは表2にまとめられている。

表1 センチメント値

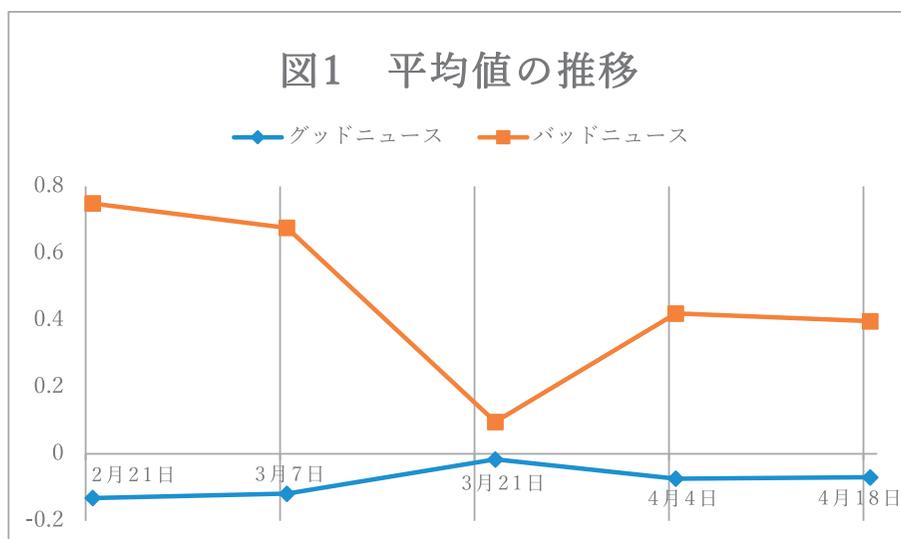
（企業名は、一部略称が使われている。並び順は、データを保存したフォルダ内の順序である）

	2月21日	3月7日	3月22日	4月4日	4月18日	合計
HOYA	-177.414	-85.2731	-103.877	-65.8793	-113.486	-545.929
KDDI	-159.506	-146.416	-116.923	-286.753	-185.535	-895.133
NTT	-272.476	-251.164	-214.151	-486.762	-321.66	-1546.21
オリエンタルランド	-55.0047	-136.219	-141.467	-143.534	-94.2923	-570.516
キーエンス	-95.4236	-242.146	-223.138	-63.0163	-238.666	-862.39
ソニーG	-70.2273	-90.1317	-85.1887	-109.119	-76.3926	-431.059
ソフトバンク G	-68.9138	-113.752	-97.0303	-109.979	-188.264	-577.939
ダイキン工業	-303.47	-272.339	-239.279	-193.71	-181.281	-1190.08
デンソー	-99.2455	-85.8937	-179.959	-56.4647	-60.974	-482.537
トヨタ	-120.643	-121.94	-175.656	-136.636	-152.35	-707.225
ホンダ	-273.76	-233.49	-887.934	-848.621	-656.463	-2900.27
リクルート HD	-127.02	-90.7578	-202.234	-34.4189	-207.532	-661.963
三井住友 FG	-136.264	-96.7432	-114.24	-227.04	-96.7553	-671.042
三菱 UFJFG	-114.699	-96.0613	-145.03	-108.112	-95.058	-558.961
信越化学工業	-76.8801	-132.794	-189.369	-129.025	-122.449	-650.518
村田製作所	-90.6702	-61.6601	-110.492	-57.5475	-49.4737	-369.843
東京エレクトロン	-61.3229	-139.352	-64.481	-67.2467	-76.604	-409.007
日本電産	-56.7073	-66.7273	-260.646	-134.473	-87.145	-605.699
日立製作所	-177.393	-105.619	-269.901	-283.88	-144.567	-981.359
任天堂	-71.2474	-68.6826	-87.3604	-35.2135	-107.526	-370.03

表2 センチメント値（標準化データ）

	2月21日	3月7日	3月22日	4月4日	4月18日	合計
HOYA	-0.63441	0.730875	0.53867	0.60374	0.390725	0.511004
KDDI	-0.39268	-0.2284	0.4619	-0.57643	0.083962	-0.09208
NTT	-1.91758	-1.8718	-0.11024	-1.64512	-0.49562	-1.21651
オリエンタルランド	1.017898	-0.06842	0.317474	0.188818	0.472444	0.468541
キーエンス	0.472313	-1.73031	-0.16312	0.619038	-0.14226	-0.03553
ソニーG	0.812419	0.654648	0.648642	0.372702	0.548656	0.709386
ソフトバンク G	0.830149	0.284067	0.57896	0.368107	0.072342	0.455722
ダイキン工業	-2.33594	-2.20401	-0.2581	-0.07928	0.102071	-0.60146
デンソー	0.420725	0.721138	0.090966	0.654044	0.614304	0.620483
トヨタ	0.131895	0.155602	0.116289	0.225674	0.225254	0.232443
ホンダ	-1.93492	-1.59451	-4.0751	-3.57861	-1.92111	-3.55499
リクルート HD	0.045817	0.644825	-0.04011	0.771839	-3.62287	-1.15498
三井住友 FG	-0.07895	0.55092	0.477691	-0.25737	0.461958	0.294931
三菱 UFJFG	0.21213	0.561618	0.296506	0.37808	0.469184	0.488498
信越化学工業	0.722618	-0.01469	0.035593	0.266338	0.35256	0.330376
村田製作所	0.536475	1.101341	0.499747	0.648259	0.663268	0.815108
東京エレクトロン	0.932613	-0.11757	0.770496	0.596434	0.547756	0.747471
日本電産	0.994915	1.021841	-0.38384	0.237232	0.502875	0.40778
日立製作所	-0.63412	0.411665	-0.4383	-0.56108	0.258391	-0.24099
任天堂	0.79865	0.991164	0.635863	0.767593	0.416101	0.814786

グッドニュースの企業群は、直前の利益予想額よりも本決算の利益額が多かった企業で、バッドニュースの企業群は、直前の利益予想額よりも本決算の利益が少なかった企業とすることを考えていたが、後になってソニーグループとソフトバンクグループの2社が利益予想の数値を公表していないことがわかった。このため、この2社については、前年度より増益ならばグッドニュース、減益ならばバッドニュースと考えることにした。この2社については、ソニーグループが増益で、ソフトバンクグループが減益であった。その他の企業については、グッドニュースとなる企業が多く、デンソーと日本電産の2社だけがバッドニュースとなった。したがってバッドニュースの企業群は、ソフトバンクを入れて3社ということになる。



グッドニュースの企業群とバッドニュースの企業群ごとの標準化されたデータの平均値の推移は、図1のようになる。バッドニュースの企業群の値は一貫してプラスである（標準化しているためグッドニュースの企業群の値は一貫してマイナスになる）。このように、グッドニュースの企業のセンチメントがポジティブ寄りで推移するという当初想定していたようなグラフは得られなかった。

グッドニュースの企業群とバッドニュースの企業群とで平均値に違いがあるかどうか、標準化前のデータを使ってt検定を試してみることにした。結果として、2月21日だけバッドニュースの企業群の平均値がグッドニュースの企業群の平均値よりも5%の水準で有意に高かった（マイナス値が小さかった）。その他の日のデータと合計値については、両企業群で平均値が有意に異なるということにはなかった。データ数も少ないので、バッドニュースの企業群の平均値がグッドニュースの企業群の平均値よりも高いと明確にいうことまではできない。

前述のDas and Chenの研究で、統合センチメント指標と株価指数は同じような動きをすることが観察されたが、各企業のセンチメントと株価の関係はあまり強くないことが明らかになっている。本研究は、各企業のセンチメントと将来的な会計情報との関係を問題にしているので、その関係があまり明確でないのは、Das and Chenの研究とも通じるところがあると思われる。

## V おわりに

本研究から得られた結果は、当初想定していたものとは違って、ある企業のグッドニュースは、その企業についての投稿のセンチメントをポジティブに寄せ、逆にバッドニュースは、その企業についての投稿のセンチメントをネガティブに寄せる

というようなことは確認できなかった。

世間全般のセンチメントと株価との関係については、Bollen ,Mao and Zengの研究でもセンチメントがポジティブだと全体的な株価水準が上昇する傾向があることが示されている。全体的な株価水準が上昇する場合、世間全般の雰囲気がポジティブになっているのではないかというのは、実感とも合致するところがある。一方、ある企業のグッドニュースを察知した者がいたとして、その企業の掲示板にポジティブな投稿をするであろうか。もう既にその企業の株を所有している株主ならば、それも考えられる。しかし一般の投資家の場合、ある企業のグッドニュースを察知した場合、安く株を買うためにネガティブな投稿（売り煽り）をすることも十分に考えられる。もしこのような傾向が認められるならば、本研究の結果もその影響を受けていると考えられなくもない。

しかし本研究はやってみることに重点が置かれているので、十分な証拠が得られているとはいいがたい。したがって上記のように、ある企業のグッドニュースは、その企業についての投稿のセンチメントをネガティブ（売り煽り）に寄せ、逆にバッドニュースは、その企業についての投稿のセンチメントをポジティブ（買い煽り）に寄せる傾向があることを検証しようとするのであれば、企業数や投稿データを収集する日数について、もっと多くする必要があるであろう。また、センチメント分析を行うために用いられるポジティブな単語、ネガティブな単語をリスト化した極性辞書について、今回は既成のものを利用させてもらった。しかし、企業の掲示板の投稿には独特の言い回しなどがある可能性もある。このため企業の掲示板の投稿に適した極性辞書を作成することができれば、より適切な検証が行えると考えられる。

企業の掲示板の投稿によって、将来のグッドニュース、バッドニュースを識別できるかということだけを問題にするならば、他の機械学習の手法を用いて分析することもできる<sup>13</sup>。ただその場合には、センチメントがポジティブなのかネガティブなのかということについてはわからなくなってしまうと考えられる。

（なかむら あきよし・高崎経済大学経済学部教授）

---

13 例えば、サポートベクターマシンやランダムフォレストなどが考えられる。