

株価データに基づく新聞記事の評価

Evaluating Newspaper Articles based on Stock Price Change

張 へい*1
He Zhang

松原 茂樹*2
Shigeki Matsubara

*1 名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University

*2 名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University

This paper describes evaluation of newspaper articles from the point of view of stock price change. The relation between phrases in the articles and stock price change is calculated by using the previous newspaper articles and stock prices. The influence of recent newspaper articles into stock price change is estimated based on the relations. In the experiment, 73,333 articles was evaluated by using the phrases appeared in 209,553 past articles. The correlation has been observed from a comparison of the estimated values with the actual stock price.

1. はじめに

近年、オンライン株取引の普及に伴い、個人投資家が急増している。機関投資家と異なり、個人投資家が利用できる情報の種類には限界があり、そのような状況下では、新聞記事は投資判断における重要な情報源となる。しかし、個人投資家が、関連するすべての記事を参照することは容易ではない。関心のある銘柄について、その株価変動に関連する記事を選択的に抽出し提供できれば、個人投資家にとって有用性が高い。

本稿では、株価変動の観点に基づく新聞記事の評価について述べる。過去の記事データと株価データを用いて記事に含まれる語句の出現と株価変動との関連を計算し、それに基づいて新しい記事内容の株価変動への影響を推測する。実験では、209,553 記事 (3 年分) に出現した語句を用いて 73,333 記事 (1 年分) を評価した。実際の株価変動データと比較した結果、両者の間の相関性を確認することができた。

2. ニュース記事と株価

2.1 株価変動と記事内容の関連性

株価は企業価値を表す指標の一つであり、企業の業績や財務等の状況、また、社会情勢の影響により変動する。一方、新聞は、上述した株価変動の要因となる事柄を広範に提供するメディアであり、記事内容と株価変動が関連することが予想される。また、ニュース報道を投資判断の材料として利用する投資家が多いことから、客観的にみて記事内容が株価変動に影響を与える可能性が高い。

典型例として、小林製薬株 (東証 1 部上場) の 2006 年 1 月下旬の株価変動を図 1 に示す。横軸は日付を、縦軸は各日の株価の終値を示す。1 月 23 日に比べ、1 月 24 日の株価が大幅に上昇している。一方、1 月 24 日の日本経済新聞の朝刊に掲載された小林製薬の関連記事を図 2 に示す。記事は、小林製薬の業績に関するものであり、株価変動と記事内容との関連性を示唆する事例となっている。

記事内容には、「経常利益が前年同期比 9% 増」「販売増加などで売上が伸びた」「製造コスト削減が進んだ」など、株価上昇の要因となる語句が現れている。本研究では、株価変動に直結する内容が記されている新聞記事には、上述のような語句が存在すると考え、過去の実績データから株価変動の要因となりやすさを語句の種類ごとに算出し、それをもとに新聞記事の

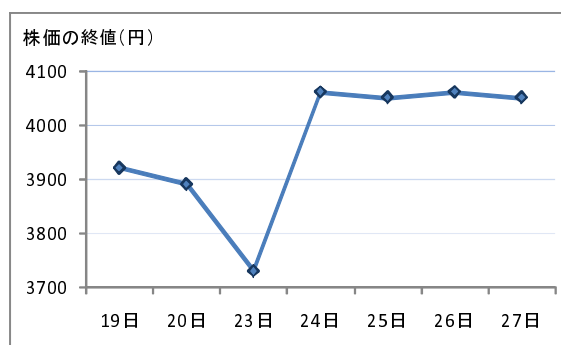


図 1: 2006 年 1 月下旬の小林製薬株の株価変動

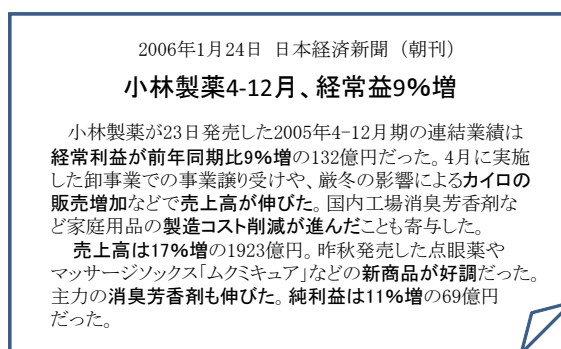


図 2: 2006 年 1 月 24 日の小林製薬の関連記事

内容を評価する。

2.2 関連研究

株価データと新聞記事の関連性については、実データを用いた実験がこれまでにいくつか行われている。小川らは、新聞記事をテーマ別に分類し、株価変動との関連性の高いテーマを判定する実験を実施している [小川 2001]。また、Koppel ら、ならびに、Mittermayer は、ある企業に関連する記事内容が、その株価の上昇要因または下降要因のいずれであるかにより、記事を分類する手法を提案している [Koppel 2004, Mittermayer 2004, Mittermayer 2006]。さらに、Lavernko らは、Web 上のニュース記事を用いて株価変動を予測するシステム AAnalyst を開

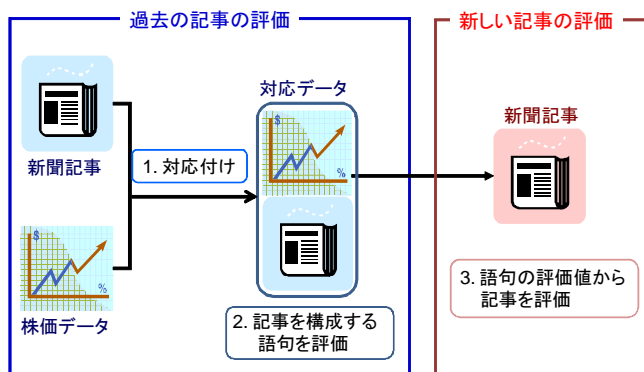


図 3: 新聞記事評価の流れ

発している [Lavernko 2001]。この他にも、株価の観点から新聞記事を分類する試みがある [Gidofalvi 2004, 高橋 2007]。

しかしながら、これらはいずれも、記事内容の特徴付けをキーワードに基づいて実行しており、記事の内容を十分に捉えられていない。また、実験については、限定された銘柄に関連する少数の記事を対象としているに過ぎず、結果の汎用性を示すには不十分である。一方、本研究では、自然言語の構文解析に基づく記事分析を行っており、様々な種類の語句を用いて記事内容を特徴付ける。また実験では、2400以上の銘柄を対象に73,333記事の評価しており、データ規模の点でも優位性が高い。

3. 株価データを用いた新聞記事の評価

本手法では、新聞記事内容を過去のデータを用いて分析し、株価変動への影響を評価する。評価の流れを図3に示す。すなわち、以下に示す手順で評価する。

1. 過去の記事データと株価変動データを対応付ける。
2. 株価データを用いて過去の記事内容の株価変動への影響を算出する。
3. 過去のデータをもとに、新しい新聞記事の株価変動への影響を評価する。

3.1 新聞記事と株価変動の対応付け

株価変動とは、企業の業績や事業等に対する市場の評価を反映したものである。そこで本研究では、株価変動率を用いて過去の新聞記事内容の株価への影響（以下、記事評価値）を計算する。

まず、日付 i における銘柄 s の株価を $p_i(s)$ とするとき、日付 i から j までの期間の株価変動率 $r_{ij}(s)$ を、

$$r_{ij}(s) = \frac{p_j(s) - p_i(s)}{p_i(s)} \quad (1)$$

で計算する。ただし、市場全体の変動の影響を排除するために、個別銘柄の評価には相対的株価変動率 $r'_{ij}(s)$ を使用する。すなわち、

$$r'_{ij}(s) = r_{ij}(s) - R_{ij} \quad (2)$$

を用いる。ここで、 R_{ij} は基準株価の変動率であり、市場全体の変動状況を示す指標を用いる。以下では、相対的株価変動率を単に株価変動率という。

記事評価値は、記事に出現するすべての銘柄の株価変動率の平均値とする。すなわち、日付 i の大引けから日付 $i+1$ の寄り付きまでに発行された新聞の記事 a に l 個の銘柄名 s_1, \dots, s_l が出現したとき、 a の記事評価値 $score_p(a)$ は、各銘柄の日付 i から j までの株価変動率を用いて、

$$score_p(a) = \frac{1}{l} \sum_{k=1}^l r'_{ij}(s_k) \quad (3)$$

で算出する。

3.2 記事の評価

過去の新聞記事に対する記事評価値を用いて新しい記事の記事評価値を推定する。記事の評価には、記事を構成する語句の評価値（以下、語句評価値）を用いる。

語句評価値は、その語句が現れる全ての過去記事の記事評価値から算出する。すなわち、語句 d が m 個の過去記事 a_1, \dots, a_m に出現したとき、 d の語句評価値 $score_q(d)$ を、

$$score_q(d) = \frac{1}{m} \sum_{k=1}^m score_p(a_k) \quad (4)$$

で計算する。

新しい記事の記事評価値の推定は、その記事に出現する語句の語句評価値の平均値の計算による。すなわち、ある新しい記事 a に n 個の語句 d_1, \dots, d_n が出現するとき、 a の記事評価値 $score_r(a)$ は、

$$score_r(a) = \frac{1}{n} \sum_{k=1}^n score_q(d_k) \quad (5)$$

で計算する。

3.3 新聞記事を構成する語句

記事を構成する語句の種類としていくつか考えられるが、本研究では依存関係 (dependency relation) を使用する。ここで依存関係とは、文構成素間の修飾・被修飾などの意味的関係をいう。すなわち、依存関係にある文構成素対を語句として使用する。本研究では、そのような文構成素対として以下の2種類を採用した。

- 文節対: 日本語係り受け構造における係り文節と受け文節の対。
- 単語対: 係り文節と受け文節の各々の主辞 (head) の対、あるいは、各文節内で依存関係にある単語対。

ただし、本研究では、隣接する単語対は依存関係にあるとみなして文節内の依存関係を抽出する。

言語表現「製造コスト削減が進んだ」から抽出される文節対、及び、単語対を図4に示す。矢印の始端と終端の対が依存関係を表している。ただし本研究では、単語間依存関係の対象を助動詞、記号以外とし、動詞については原形を用いる。また、係り文節内の助詞の種類（は、が、の、を、に、へ）により依存関係の種類を定め、主辞間の単語対に導入し、その場合、「依存元」「依存先」「依存の種類」の3項組で表現する。

4. 実験

4.1 実験の概要

新聞記事と株価変動の関連性を調べるために、実データを用いた分析実験を実施した。実験は、2003年から2005年まで

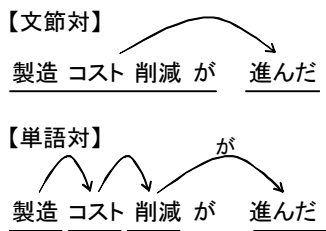


図 4: 依存関係の例

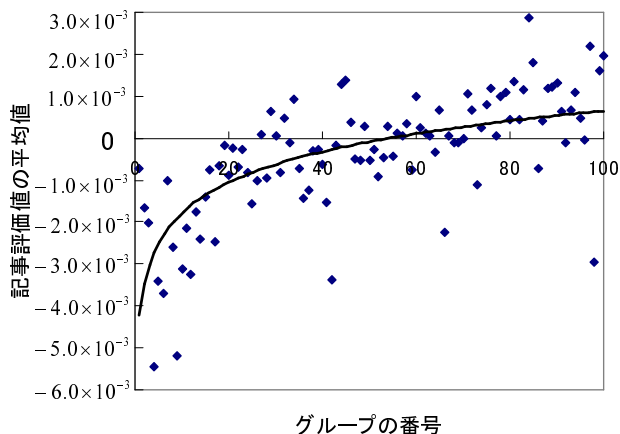


図 5: 記事評価値と株価変動率の関係 (単語対)

の 3 年間の 209,553 記事ならびに株価データを用いて、2006 年の 1 年間の 73,333 記事の評価値を推定し、それを実際の株価変動データと照合することにより行った。新聞記事として、日本経済新聞社が発行する「日本経済新聞」「日経金融新聞」「日経産業新聞」の 3 紙を使用した。株価データとして、東京証券取引所に上場している全銘柄 (東証 1 部, 東証 2 部, 東証 マザーズ) を対象とした。2006 年には、2,416 銘柄が上場していた。

銘柄の株価変動率の算出では、単位期間を 1 日 (すなわち、 $j = i + 1$) とし、株価には各日の終値を用いた。また、基準株価には、日経平均株価を使用した (式 (2) 参照)。新聞記事と株価変動率の対応付けについては、日本経済新聞朝刊、日経金融新聞、日経産業新聞に対しては発刊当日の株価変動率を、日本経済新聞夕刊に対しては発刊翌日の株価変動率を使用した。

実験における評価値推定のための語句として、文節対、及び、単語対に加え、単語を単一で用いる方法も比較のために実施した。なお、形態素解析には chasen^{*1} を、係り受け解析には cabocha^{*2} を使用した。

4.2 実験結果

本手法により推定した記事評価値と、記事に出現する銘柄の株価変動率との相関を調べた。すなわち、2006 年の全記事 (73,333 記事) を、推定した記事評価値の昇順ソートにより、100 グループに等分割し (グループ 1 が最も推定値の低い記事、グループ 100 は最も高い記事から構成)、記事グループとその

表 1: 記事評価値と株価変動率の関係 (語句の種類ごとの相関)

語句の種類	単語	文節対	単語対
順位相関係数	0.52	0.59	0.71

記事評価値平均との間の順位相関係数を算出した。

表 1 に、使用した語句の種類ごとの順位相関係数を示す。いずれを使用した場合においても、両者の間に相関があることが認められるが、単語を単独で用いる場合よりも、依存関係を使用した場合の方が相関が高く、また、依存関係としては単語対を用いた場合に相関が最も高く、相関係数は 0.71 となった。単語間依存関係を用いた場合における、記事グループ番号とグループごとの記事評価値の平均をプロットしたグラフを図 5 に示す。このグラフからも両者の高い相関性を確認できる。

5. おわりに

本稿では、新聞記事内容と株価変動との関連性の分析について述べた。記事を構成する言語表現に基づいて算出した新聞記事の評価値と、実際の株価データとを比較した。3 年分の記事に出現した言語表現を用いて 1 年分の記事を評価し、株価変動データと比較した結果、両者の間の相関性を確認することができた。今後は、記事内容をより特徴付けるために、記事を構成する言語表現の種類について検討する予定である。

参考文献

- [小川 2001] 小川知也, 渡部勇: 株価データと新聞記事からのマイニング, 情報処理学会研究報告, NL142-19, pp.137-144 (2001).
- [Koppel 2004] Koppel, M. and Shtrimberg, I.: Good News or Bad News? Let the Market Decide, *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pp.86-88 (2004).
- [Mittermayer 2004] Mittermayer, M. A.: Forecasting Intraday Stock Price Trends with Text Mining Techniques, *Proceedings of 37th Hawaii International Conference on System Sciences*, pp.64-73 (2004).
- [Mittermayer 2006] Mittermayer, M. A., Gerhard, F. K.: Text Mining Systems for Market Response to News: A Survey, <http://www.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-184.pdf> (2006).
- [Lavernko 2001] Lavernko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., and Allan, J.: Mining of Concurrent Text and Time Series, *Proceedings of KDD Conference Text Mining Workshop*, pp.37-44 (2001).
- [Gidofalvi 2004] Gidofalvi, G.: Using News Articles to Predict Stock Price Movements, <http://citeseer.ist.psu.edu/gidofalvi01using.html> (2004).
- [高橋 2007] 高橋悟, 高橋大志, 津田和彦: ヘッドラインニュースに対する株価の反応について, <http://www.iser.osaka-u.ac.jp/rcbe/6thworkshop/presentationHP.pdf/SatoruTakahashi070210.pdf> (2007).

*1 <http://chasen.naist.jp/>

*2 <http://chasen.org/taku/software/cabocha/>