

# 複数のメディアのセンチメントの推移と株価変動の関係性の分析

西開地 晃司<sup>†</sup> 豊田 哲也<sup>†</sup> 大原 剛三<sup>†</sup>

<sup>†</sup>青山学院大学 理工学部

## 1. はじめに

株式投資において判断材料になる情報は、株価の推移や経済指標、IR(Investor Relations)や SNS(Social Networking Service)などがある。近年、これらの情報を分析して得られる評価表現や感情スコア（以下、センチメントと表現）を用いて株価変動を予測する研究が行われている[1, 2]。それらの研究では、1つのメディアを対象に分析を行っている。投資に関係のある情報は複数のメディア上で掲載されているため、複数あるメディアのうちあるメディア上のセンチメントと投資対象銘柄の株価の相関が分かれば、投資の際に重視する情報源の判断の指針となる。そのため本研究では、複数のメディアから収集したテキストを使用して個別銘柄に対するセンチメントと株価の関係性を明らかにする。また、テキストから抽出したセンチメントと個別銘柄の株価の前日比（以下、前日比と表現）を用いて、株価騰落予測モデルの構築を試みる。

## 2. 提案手法

### 2.1 概要

本システムは、図1のように大きく「データ収集」「感情分析」「重回帰分析」の3つのステップにより構成される。本研究で対象とするメディアは Twitter<sup>1</sup>, Yahoo! Japan が運営する株式掲示板 Textream<sup>2</sup>, 産経新ニュースの経済コラム<sup>3</sup>とする。

### 2.2 データ収集

Twitter が提供する Streaming API を用いて、対象期間中のツイートを収集した。取得したツイートデータは「ツイート文」「投稿日時」「ユーザ名」とした。同一ユーザが同じツイートを複数回ツイートした場合は、1ツイートのみを使用した。産経ニュースと Textream のテキストは、Web ページの情報を自動で取得するクロ

Cross-media Analysis of the Relationship between Consumer Sentiment and Stock Prices

Koji Nishikaichi<sup>†</sup>, Tetsuya Toyota<sup>†</sup> and Kouzou Ohara<sup>†</sup>

<sup>†</sup>College of Science and Engineering, Aoyama Gakuin University

<sup>\*1</sup> <https://twitter.com/>

<sup>\*2</sup> <http://textream.yahoo.co.jp/>

<sup>\*3</sup> <http://www.sankei.com/economy/economy.html>

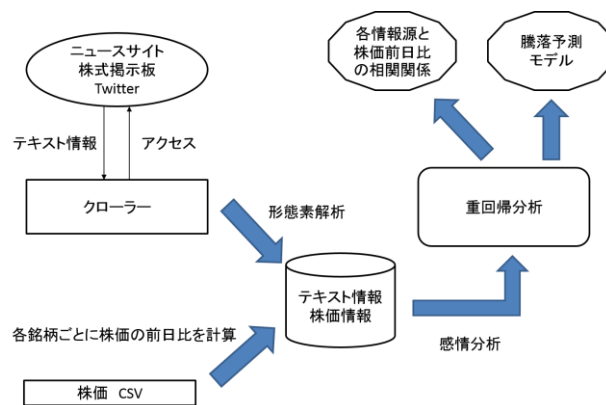


図1 提案システムの概要

ーラを作成して集計した。産経ニュースは「投稿日時」「ニュースタイトル」「ニュース記事」、Textream は「投稿日時」「投稿テキスト」「投稿者名」「企業名」を取得した。株価情報に関しては、株価データサイト k-db.com<sup>4</sup>から個別銘柄の時系列 CSV データを取得し、データベースに格納する。その際に当日の株価終値を前日の株価終値で割った前日比が 1 より大きければ 1, 1 であれば 0, 1 未満であれば -1 というように離散化し格納する。Twitter と産経ニュースは、Textream の掲示板のように企業ごとにテキストデータが分けられていない。そのため、収集したテキストに形態素解析エンジン MeCab を使用して、抽出した形態素の品詞情報に「固有名詞」「組織」を含む名詞を格納する。また、次のステップで係り受け解析を行うために、テキストに存在する改行文字を句点に置換をしたものデータベースに格納する。

### 2.3 感情分析

データベースに格納したテキストのうち対象企業名を含むものに対して感情分析を行い、それぞれにセンチメントを付与する。その際、「ない」や「ません」のような否定語を含む文に関しては、係り受け解析により単語の極性を反転させる。ただし、二重否定の場合は極性の反転を行わない。本研究では、センチメントの計算には、乾・岡崎研究室が公開している日本語評価極性辞書（用言編・名詞編）[3, 4]を利用

<sup>\*4</sup> <http://k-db.com/>

し、係り受け解析には日本語構文解析器 CaboCha を用いた。

### 2.4 重回帰分析

算出したセンチメントを企業ごとに集計し、日毎の平均を算出する。説明変数は情報源ごとに集計したセンチメント、目的変数は対象銘柄の株価の前日比として重回帰分析を行う。

## 3. 評価実験

### 3.1 実験設定

本実験では、2016年8月23日から2016年12月16日の期間に各メディアから収集したテキストを対象に各メディアと対象銘柄間の関係性を分析した。センチメントは、「0時～24時」「9時～15時」「前日15時～9時」「15時～翌日9時」の4種類の単位で集計した。対象企業数は表1に示す24社とした。

### 3.2 実験結果と考察

トヨタ自動車、日産自動車、ホンダを対象に8月から11月までを訓練データ、12月をテストデータとして重回帰分析を行った。1日後の前日比、2日後の前日比、3日後の前日比を目的変数として重回帰分析の結果を表2、表3、表4に示す。もっとも妥当であると思われるモデルは、ホンダの1日後の前日比を用いたモデルである。このモデルは、有意確率は0.05よりも低いため95%以上の確率で妥当であると思われる。この際Twitterの有意確率が0.01よりも低いため、Twitterがホンダの情報源として最も重視すべきメディアであるといえる。一方、日産自動車とトヨタ自動車は表2、表3、表4を見ても、モデルの有意水準が5%を満たすものがなかった。

### 4. おわりに

本研究では、複数のメディアのテキストからセンチメントを抽出し、対象銘柄の株価の前日比の相関を分析した。実験結果から、ホンダは最も要視すべき情報源が見つかったが、トヨタ自動車と日産自動車は見つからなかった。ホンダは、Twitterの投稿記事数が最も高く、重視すべき情報源としてもTwitterという結果が得られたが、トヨタ自動車と日産自動車に関してはそのような傾向が見られなかった。

表1 実験に使用した対象企業

企業名
トヨタ自動車、日産自動車、ホンダ、ソニー、東芝、日立製作所、パナソニック、マツダ、ソフトバンクグループ、スズキ、電通、富士通、昭和シェル、東京電力ホールディングス、三菱商事、三菱電機、富士重工業、三菱電機、三井物産、関西電力、ヤフー、日本郵船、資生堂、東武鉄道、川崎重工業

表2 1日後の前日比を対象

企業名	集計区分	偏回帰係数 最も高い情報源	T検定 P値	F検定 P値	予測精度
トヨタ自動車	0時～24時	Textream -0.2775	0.0255	0.0763	0.55
日産自動車	9時～15時	Textream -0.1406	0.268	0.4946	0.55
ホンダ	0時～24時	Twitter -0.3352	0.0080	0.0415	0.55

表3 2日後の前日比を対象

企業名	集計区分	偏回帰係数 最も高い情報源	T検定 P値	F検定 P値	予測精度
トヨタ自動車	15時～翌日9時	Textream -0.2900	0.0216	0.1284	0.50
日産自動車	0時～24時	Twitter -0.2090	0.0995	0.3385	0.60
ホンダ	0時～24時	Textream -0.2739	0.0309	0.0913	0.55

表4 3日後の前日比を対象

企業名	集計区分	偏回帰係数 最も高い情報源	T検定 P値	F検定 P値	予測精度
トヨタ自動車	0時～24時	Textream 0.1957	0.1301	0.4431	0.44
日産自動車	0時～24時	Twitter 0.2613	0.0400	0.2143	0.44
ホンダ	0時～24時	Twitter -0.3352	0.0080	0.0809	0.55

### 参考文献

- [1] 前川浩基, 中原考信, 岡田克彦, 羽室行信: 大規模ニュース記事からの極性付き評価表現の抽出と株価収益率の予測, オペレーションズ・リサーチ: 経営の科学, Vol. 58, No. 5, pp. 281-288, 2013.
- [2] 坪内孝太, 山下達雄: 株価掲示板情報の感情解析と株価との相関の研究, 2015年度人工知能学会全国大会講演集, 1J5-OS-13b-2in, 2015.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 3, pp. 203-222, 2005.
- [4] 東山昌彦, 乾健太郎, 松本裕治: 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第14回年次大会論文集, pp. 584-587, 2008.