

# 株価情報と新聞記事を用いた企業の評判情報の抽出に関する研究

中村健二† 田中成典‡ 吉村智史† 細島啓史† 北野光一† 小谷章‡

関西大学大学院総合情報学研究科† 関西大学総合情報学部‡

## 1. はじめに

近年、ある対象の評判に関わる情報のうち Web 上に存在するものは様々な用途[1]で使われており、評判情報として注目が高まっている。その中でも新聞記事中の評判情報は、情報そのものの信頼性が高いため、消費者が商品を購入する際の商品提供企業の評判情報や、株式売買における企業の決定など、人々の企業に対する行動方針を決定付ける情報として利用[2]される。しかし、新聞記事から複数の企業に関する情報を手作業で集めるためには多大な時間がかかる。そのため、大量の新聞記事から企業の評判に関わる情報を自動抽出することが望まれている。新聞記事から企業の評判に関わる情報を自動抽出する既存研究として、株価の終値情報を手がかりとして、新聞記事を読者に良い印象を与える新聞記事と悪い印象を与える新聞記事に分ける研究[3]や、株価の変動原因を新聞記事から抽出する研究[4][5]がある。しかし、いずれの研究も、株価の変動時期を決定する上で、求める時期とその1日後についての終値しか比較しておらず、求める時期以前からの株価変動を考慮していない。そのため、既存研究の手法では、株価が急激に変動した際に誤認識する可能性があり、企業の評判情報を適切に抽出できない場合がある。そこで、本研究では、求める時期以前と以降の両方からの株価変動を考慮した株価の終値情報を手がかりとして、新聞記事から良い評判と悪い評判に分類した企業の評判情報を抽出する手法を提案し、その有用性を検証する。

## 2. システムの概要

本研究では、対象とする企業の株価情報と新聞記事を基に企業の評判情報を抽出することを

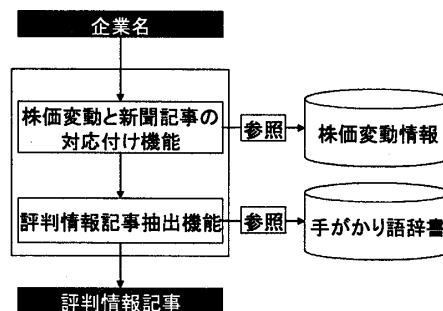


図1 システムの流れ

目的とする。本システムは、図1に示すように企業名を入力とし、評判情報記事を出力とする。本システムは、1) 株価変動と新聞記事の対応付け機能、2) 評判情報記事抽出機能の2つの機能で構成される。

### 2.1 株価変動と新聞記事の対応付け機能

本機能では、入力企業についての株価変動情報を株価変動情報データベースから取得し、新聞記事と対応付ける。株価変動情報データベースには、株価の変動情報を Web 上から取得し、正規化を施した株価変動情報を格納する。株価変動情報には、変動状態として「上昇」、「下降」と「平坦」の3種類がある。株価変動情報の取得後は、入力企業名を含む新聞記事を全て収集し、新聞記事の掲載時期前後の変動状態に対して、新聞記事を対応付ける。抽出した新聞記事は、新聞記事の掲載時期前後で変動状態が好転した新聞記事をポジティブ情報が内在する新聞記事とし、新聞記事の掲載時期前後で変動状態が悪化した新聞記事をネガティブ情報が内在する新聞記事とする。

### 2.2 評判情報記事抽出機能

本機能では、ポジティブ情報またはネガティブ情報が内在するとされた全ての新聞記事に対して、手がかり語に基づいた評判情報の抽出を行う。対象とする新聞記事のうち、手がかり語を含む新聞記事を評判情報記事とし、手がかり語を含む一文を評判の原因部分とする。なお、手がかり語は、手作業で取得した少量の正解単語を基に、新聞記事から自動学習することによって事前に生成した手がかり語辞書データベースから取得する。

Research for Extracting Popular Information of Company using Stock Information and News Articles

†Kenji Nakamura, Satoshi Yoshimura, Hirofumi Hosohata, Koichi Kitano

Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

‡Shigenori Tanaka, Akira Odani

Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

### 3. システムの実証実験と考察

本システムの有用性を実証するために評判情報の抽出精度に関する実験を行った。

#### 3.1 実証実験

評判情報抽出の実験では、手がかり語を使用した新聞記事からの評判情報抽出システムについて、提案手法と既存手法[3]との比較実験を行った。実験用の新聞記事として、2002年11月1日から5年間の新聞記事のうち、入力企業名を含む新聞記事をヨミダス文書館から取得した。また、実験用の終値情報として、入力企業に関する同期間の終値情報をYahoo!ファイナンスから取得した。実験では、企業の名称を入力とし、実験用のデータから自動抽出した評判情報について評価を行った。評価は、自動抽出した評判情報と正解データとの比較によって行った。なお、正解データとして、実験用の新聞記事から目視で分類したポジティブ情報とネガティブ情報を使用した。また、評価指標にはF値を使用した。F値は、適合率と再現率の調和平均から算出する評価指標であり、情報検索の精度に関する一般的な尺度として用いる。適合率とは、評判情報と判定された基準のうち何件が実際に評判情報かを示す基準である。また、再現率とは、全ての新聞記事に含まれる実際の評判情報のうち何件が取得できたかを示す基準である。

#### 3.2 結果と考察

本実験では、本システムで抽出したポジティブ情報とネガティブ情報それぞれについて、既存システムとの比較を行った。実験結果を表1、表2と表3に示す。表1より、評判情報記事の取得件数について優れた結果となり、表2と表3より、再現率についても優れた結果となったことが分かった。これらは、株価が急激に変動した際に発生した誤認識による検索漏れが改善された結果であると考えられる。しかし、適合率は既存システムと同程度か、または低い値となることが認められた。これは、新聞記事の取得範囲が広まったことにより、ノイズが増えたためであると考えられる。F値及び取得件数から、本提案手法の方が有意であることが示されるが、適合率について改善の余地があると考えられる。

#### 4. おわりに

本研究では、株価の変動情報に基づいて新聞記事から入力企業の評判情報を抽出し、ポジティブ情報とネガティブ情報に分類して提示した。そして、実証実験の結果から本研究の有用性を実証した。このことから、人々が行動方針を決定付ける上で有益な情報を整理できたと考える。

表1 評判情報記事の取得件数

|       |         | 取得件数 |
|-------|---------|------|
| 既存手法  | ポジティブ情報 | 18   |
|       | ネガティブ情報 | 9    |
| 本提案手法 | ポジティブ情報 | 29   |
|       | ネガティブ情報 | 21   |

表2 ポジティブ情報抽出の実験結果

|       | 再現率  | 適合率  | F値   |
|-------|------|------|------|
| 既存手法  | 0.60 | 0.36 | 0.45 |
| 本提案手法 | 0.96 | 0.48 | 0.64 |

表3 ネガティブ情報抽出の実験結果

|       | 再現率  | 適合率  | F値   |
|-------|------|------|------|
| 既存手法  | 0.30 | 0.90 | 0.45 |
| 本提案手法 | 0.70 | 0.35 | 0.46 |

今後は、入力企業の株価変動と日経平均株価や東証株価指数など、業界の平均を示す株価変動との比較に基づいた企業の業績による株価の変動時期推定や、株価の急変だけではなく様々な変動パターンに基づいた評判情報の出現位置推定によって誤認識を抑え、適合率の向上を図る予定である。

#### 参考文献

- [1] 立石健二, 石黒義英, 福島俊一: インターネットからの評判情報検索, 人工知能学会誌, 人工知能学会, Vol.19, No.3, pp.317-323, 2004.5.
- [2] Fogg, B. J., Kameda, T., Boyd, J., Marshall, J., Sethi, R., Sockol, M., and Trowbridge, T.: Investigating What Makes Web Sites Credible Today, A Research Report by the Stanford Persuasive Technology Lab & Makovsky & Company, Stanford University, 2002.5.
- [3] 酒井浩之, 増山繁: 経済新聞記事内容の個々の企業におけるインパクトの判定, 情報学基礎研究会研究報告, 情報処理学会, Vol.2006, No.94, pp.43-50, 2006.9.
- [4] 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, 情報学基礎研究会研究報告, 情報処理学会, Vol.2001, No.20, pp.137-144, 2001.3.
- [5] Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., and Allan, J.: Language Models for Financial News Recommendation, Proceedings of the 9th International Conference on Information and Knowledge Management, ACM SIGMIS, pp.389-396, 2000.6.