

# ニュース記事から株価が変動する企業の特定に関する研究

山本雄平<sup>†</sup> 中村健二<sup>‡</sup> 田中成典<sup>‡†</sup> 柳田尚明<sup>†</sup> 井内敦<sup>‡†</sup>

関西大学大学院総合情報学研究科<sup>†</sup> 立命館大学情報理工学部<sup>‡</sup> 関西大学総合情報学部<sup>‡†</sup>

## 1. はじめに

インターネットには、企業に関連したニュース記事が多く蓄積[1]されており、それらの情報を解析して株価変動の原因を説明、分析する研究が行われている。小林ら[2]は、日経平均株価の値動きについて、上昇や下落などのパターン別に分類し、株価変動の原因を値動きと関連付けて説明する手法を提案している。Ahmadら[3]は、ロイターの配信記事から株価に関する情報を抽出し、記事に含まれる数値情報からグラフを生成し、記事の要約と併せて提示する手法を提案している。和泉ら[4]は、金融経済月報のレポートを解析して、株価の時系列的な変動を説明するキーワードグラフを提示する手法を提案している。中村ら[5]は、新聞記事から企業の評判情報を収集し、株価変動と対応付ける研究を行っている。しかし、これらの研究では、ニュース記事に掲載されている特定の銘柄や企業に限定して分析しているため、ニュース記事に掲載されていない企業に対して、そのニュース記事の内容が影響を与えるかを解析できない。そこで、本研究では、過去のニュース記事を解析することで、ニュース記事のトピックに関連する企業を抽出する手法を提案する。これにより、ニュース記事に掲載されていない企業に対しても、株価変動の影響の有無を把握できる。

## 2. 研究の概要

本研究では、ニュース記事のトピックを考慮することで、解析対象のニュース記事が株価に影響を与える企業を抽出する手法を提案する。本システムの概要を図1に示す。入力データはニュース記事とし、出力データは、企業グルー

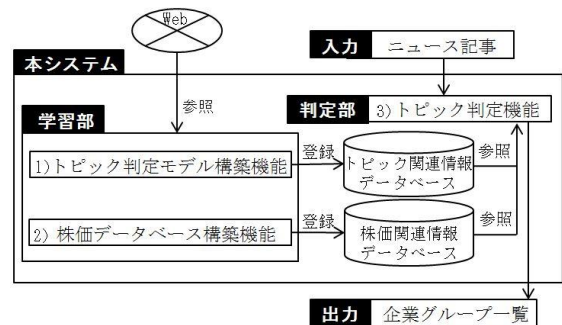


図1 本システムの概要

プー一覧とする。本システムは、1) トピック判定モデル構築機能、2) 株価データベース構築機能、3) トピック判定機能の3つの機能により構成される。

### 2. 1 トピック判定モデル作成機能

本機能では、まず、Yahoo!ニュースの経済カテゴリに含まれるトピック179件のニュース記事を抽出する。次に、ニュース記事のHTMLを解析して、記事のタイトル、配信日時、本文、記事に掲載されている企業名を抽出し、トピック毎に保存する。そして、各トピックの記事本文を形態素解析して、TF-IDF値を算出し、TF-IDF値の上位1,000件を各次元とした素性ベクトルを作成する。最後に、各トピックの素性ベクトルを入力として、SVM (Support Vector Machine) にてトピック判定モデルを構築する。

### 2. 2 株価データベース構築機能

本機能では、トピック毎に保存された各企業の株価を過去3年分取得し、データベースに蓄積する。取得する情報は、銘柄コード、日足の終値情報とする。

### 2. 3 トピック判定機能

本機能では、まず、入力されたニュース記事に対して、トピック判定モデル作成機能と同様の処理を行い、ニュース記事の素性ベクトルを作成する。次に、トピック判定モデルとニュース記事の素性ベクトルをSVMにて解析し、ニュース記事が所属するトピックを判定する。そして、判定されたトピックに関連する企業と株価変動の影響を受けた割合を算出して提示する。

Research of Method for Specifying Private Enterprise with Stock Price Effected from News Articles

<sup>†</sup> Yuhei Yamamoto, Naoaki Yanagida

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

<sup>‡</sup> Kenji Nakamura

College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Norihigashi, Kusatsu City, Shiga, 525-8577, Japan

<sup>‡†</sup> Shigenori Tanaka, Atsushi Iuchi

Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki City, Osaka 569-1095

### 3. システムの実証実験と考察

本システムの有用性を確認するために評価実験を行い、出力結果について分析した。

#### 3.1 実証実験

実証実験では、Yahoo!ニュースの経済カテゴリに登場した記事を用いて提案システムの有用性を評価する。利用したニュース記事は、12月26日から12月28日に配信された中で、企業名が登場する2件を用いる。ここで、実験に用いるニュース記事は、特定の企業に関するものではなく、複数の企業に関連するものを選択する。

実験に用いた記事は、携帯電話関連の記事(記事 No.1)とLEDの生産に関する記事(記事 No.2)である。

#### 3.2 結果と考察

本システムの実行結果を表1、入力した記事を元に抽出した関連企業名の一例を表2、過去の事例における株価への影響の有無を表3に示す。

表1は、入力したニュース記事のトピックをシステムが正しく判定できているかどうかを分析した結果である。この結果を確認すると、記事 No.1 と記事 No.2 ともに正確にトピックを判定出来ていることがわかる。また、トピックに関連しており、ニュース記事に登場していない新出企業も抽出できており、本研究の目的を達成できていることがわかる。

表2は、入力したニュース記事のトピックに関連する企業群の中から一例を取り出した結果である。記事 No.1 の関連企業名を確認すると、シャープなどの携帯電話関連の企業名が抽出できていることがわかる。また、直接的な関係はないと思われる東京電力などの企業も含まれていたが、データを確認すると、過去の記事において、東京電力がKDDIの株を売却したニュースが含まれていたことが原因であると考えられる。

表3は、表1の新出企業の中から、入力したニュース記事のトピックで株価が大きく変動した企業の変動率上位3件を示している。表3の記事 No.2 の結果を確認すると、東日本大震災の影響により、電力不足に備えて街灯のLED化が進んでいるため、LEDとの関連が深いと思われるノアや東京電力が株価変動企業として抽出できていることがわかる。

これらの実験結果より、本提案手法はニュース記事に関連する可能性のある企業を抽出でき、一定の成果を得ることができた。しかし、表3で1位になった「ランド」など、企業名と一般名詞が区別せずに抽出される問題が明らかになったため、企業名を高精度に評価する文脈解析手法や複合語用語抽出手法などを組み合

表1 本システムの実行結果

記事 No.	元のトピック	判定トピック	登場企業数	新出企業数	株価変動企業数
1	KDDI	KDDI	1件	152件	33件
2	LED(発光ダイオード)	LED(発光ダイオード)	1件	64件	10件

表2 抽出した関連企業名の一例

記事 No.	企業名
1	シャープ, サイバーエージェント, 東京電力, オリンパス, サンリオ, ドリコム, オービス
2	東芝, 富士通, ソニー, シャープ, 村田製作所, 高島屋, イオン, ツインバード工業

表3 過去の事例における株価変動の有無

記事 No.	企業名	株価変動率
1	ランド	0.93(76件)
	オリンパス	0.89(73件)
	テンボスバスターズ	0.88(72件)
2	ランド	0.91(52件)
	ノア	0.68(39件)
	東京電力	0.65(37件)

せることが必要と考えられる。

#### 4. おわりに

本研究では、ニュース記事を入力し、トピックで扱われた企業名と株価変動を考慮した関連企業名の抽出手法について検討し、実験を通じて、本提案手法の有用性を確認した。また、ニュース記事では言及されていない多数の企業名を取得できたことから、潜在的な企業間関係を考慮できたと考える。今後は、文脈や複合語を考慮した正確な企業名の評価と、企業間の株価の関連についての評価を行い、システムの改良に取り組む予定である。

#### 参考文献

- [1] 総務省：平成23年度版情報通信白書，ぎょうせい，2011.
- [2] 小林一郎，渡邊千明，奥村奈穂子：グラフとテキストの協調による知的な情報提示手法-日経平均株価テキストとグラフの提示を例にして，情報処理学会論文誌，情報処理学会，Vol.48，No.3，pp.1058-1070，2007.
- [3] Ahmad, S., Olibeira, P. and Ahmad, K.: Summarization of Multimodal Information, *Proceedings of The 4th International Conference on Language Resources and Evaluation*, LREC 2004, Vol.3, pp.1049-1052, 2004.
- [4] 和泉潔，後藤卓，松井藤五郎：テキストマイニングによる金融市場の月次動向分析，知能と複雑系研究会研究報告，情報処理学会，Vol.2009，No.16，pp.75-80，2009.
- [5] 中村健二，田中成典，吉村智史，細島啓史，北野光一，小谷章：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，全国大会講演論文集，情報処理学会，Vol.70，No.1，pp.637-638，2008.