

ニュース記事の考慮の有無による株価指数の予測結果の差に基づく 経済的影響力の推定

米田 宏生[†] 湯本 高行[‡] 磯川 悌次郎[†] 上浦 尚武[†]

兵庫県立大学 大学院工学研究科電子情報工学専攻[†] 兵庫県立大学 社会情報科学部[‡]

1 はじめに

近年、ビッグデータ分析やAIの発展により、それらの金融分野での活用が注目されている。実際にビッグデータである時系列データの分析を基に株価を予測する研究などがある。また、金融に関係するテキストとしてニュース記事がある。様々な出来事について書かれているニュース記事だが、中には企業の買収や政策の発表などの経済に影響を与え、株価変動の要因となるものもある。このような様々なニュース記事がある中で、経済に影響を与えるものを見つけることができれば、企業の経営戦略や投資判断に活用することができる。しかし、ニュース記事は日々膨大な数が配信されており、人手によって確認し、影響力を推定するのは難しい。そこで、本研究では株価指数とニュース記事を用いることでそのニュース記事のもつ影響力を推定する手法を提案する。本手法を用いることで、定量的に経済の流れを捉え、そしてニュース記事の影響力を推定することが可能である。

2 関連研究

株価を予測する研究は数多くある。テキストと時系列データを用いて株価を予測する研究として、Vargasらの研究[1]がある。これは株価指数とニュースタイトルから株価騰落を予測する研究である。しかし、株価の騰落を予測することが目的となっているため、株価自体の予測は行っていない。一方で、提案手法は、時系列データとテキストデータを組み合わせて株価予測を行い、そこからニュース記事の影響力を推定する。そのため、本手法はニュース記事の影響力を推定するのみならず、株価予測にも応用が可能な手法である。

3 提案手法

3.1 影響力の推定モデル

本手法では株価指数に基づく特徴量とニュー

ス記事を入力とし、株価指数を予測するモデルを構築し、以下の2つの入力から得られた出力の差(株価+ニュース - 株価のみ)として影響力を算出する。

- 株価指数に基づく特徴量とニュース記事を入力として与えた場合の出力(株価+ニュース)
- 株価指数に基づく特徴量はそのまま入力するが、ニュース記事に対応する入力をすべて0として与えた場合の出力(株価のみ)

影響力が正の場合は当該ニュースはポジティブな影響があり、負の場合はネガティブな影響があると考えられる。なおこのモデルの出力の株価指数は、経済への影響を推定するために業種単位で定義する。まず、TOPIX-17の企業分類に基づき、各業種上位10社組入銘柄を選択し、それら個別企業の株価を標準化する。そして純資産比を基に加重平均をとった数値を出力として扱う。

本研究で使用するモデルは図1に示すように、以下の2つを全結合層に入力することによって、株価指数を予測する。

- 株価指数に基づく特徴量を入力とするLSTM層の出力
- ニュース記事のタイトルおよび本文の埋め込み表現のそれぞれを入力とするLSTM層の出力

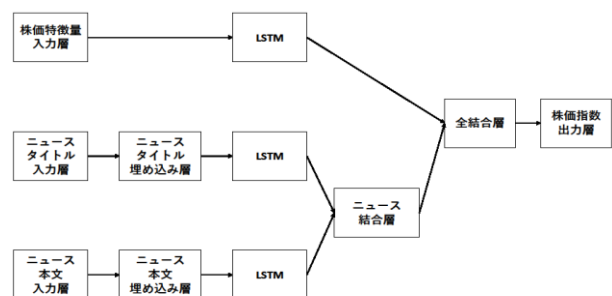


図1 モデルのニューラルネットワーク構造

3.2 株価指数に基づく予測

本研究では、ニュース記事の経済への影響を推定するために、業種単位でのモデルの構築を考える。そのため、各業種の上位10社の株価から作成した株価指数から終値の変動に加え、7つのテクニカル指標[2]を算出し、特徴量とする。テクニカル指標の算出における変数は[2]に従う。

Estimation of economic impact based on difference in predicted stock index with/without news article

[†]Kosei Yoneda, Teijiro Isokawa, Naotake Kamiura
University of Hyogo

[‡]Takayuki Yumoto, University of Hyogo

また、LSTM層への入力、過去7日間のデータを基に翌日の株価指数を予測するデータ形式をとり、ユニット数は128を設定する。

3.3 ニュース記事に基づく予測

本研究では、代表企業の株価とその企業と関連のあるニュース記事を用いることで影響力を推定していくため、ニュースに関係する企業が明確に記載されている日経速報ニュースのデータを用いる。このデータには内容を端的に示す分類語という情報がある。そこで、その情報を用いることで予め経済に影響を与えないニュースを省く。次にニュース記事を構成する要素の中から、ニュースタイトルとニュース本文を特徴量として用いる。また、タイトルと本文に含まれる名詞のみを対象にし、形態素解析器JUMAN[3]の分析結果から企業名などの固有名詞を特定し、[組織名]というように抽象的表現に置き換える。そして、埋め込み層にて名詞のみで表現されたタイトルや本文を朝日新聞単語ベクトル[4]を用いてベクトル化する。LSTM層は、ユニット数が128で、埋め込み層でベクトル化された300次元のベクトルを入力としている。最後に、ニュースタイトル層とニュース本文層での出力を結合する。

4 評価実験

4.1 実験方法

今回は17業種あるうちの、素材・化学、電力・ガス、金融の3業種を選択し、モデルを構築する。学習データとして、2014年、2016年、2017年の株価指数とニュース記事を用いる。そして、2018年1月から3月のデータを入力した際の影響力を推定する。

4.2 実験結果と考察

選択した3業種の2018年1月から3月までのデータを入力して得られた結果の例を表1に示す。まず、各業種の影響力の大きさの平均を算出すると、素材・化学は0.0232、電力・ガスは0.0198、金融は0.0240となった。素材・化学と金融は景気や金利動向に敏感に反応する業種である一方で、電力・ガスはインフラ系で安定している業種である。そのため、電力・ガスの影響力の大きさが小さくなる傾向にあるのは、業種の特徴を捉えた推定ができていたためだと考えられる。そして、表1より買収、出資、業務提携といったM&Aに関するニュースは比較的影響力が大きくでる傾向にあった。ただ、M&A関連ニュースでも関わってくる企業の経営状況などが異なるため、影響力のポジティブ・ネガティブは一概に特定できるものではないと考えられる。また、コー

ポレート・ガバナンスに関するニュースが非常に小さい影響力と推定されているが、内容が自社内で完結するものであるため、M&Aに関するニュースよりも小さくなっていると考えられる。

表1 影響力例

タイトル	影響力
東北電力、東急パワーサプライへ出資	0.0696
野村HDとLINE、金融事業の業務提携検討開始に関する基本合意書を締結	0.0451
三菱ケミカル、3Dプリンター用フィラメントメーカー蘭Dutch Filaments社を買収	-0.0336
オリックス、オマーン現地法人「Oman ORIX Leasing Company SAOG」の全株式を譲渡	-0.0305
東北電力、経営機構の見直しによるコーポレート・ガバナンスの強化について発表	0.0006

5 おわりに

本研究では、株価指数とニュース記事を用いることで、そのニュース記事の影響力を推定する手法を提案した。その結果、影響力の大小によって傾向が見られた。今後の課題として、ニュース記事の影響期間を考慮したモデルを構築することがあげられる。

謝辞

本研究はJSPS科研費JP19H04116の助成を受けたものです。

参考文献

- [1] Manuel Ramon Vargas, Beatriz S.L.P de Lima, Alexandre Evuskoff, Deep learning for stock market prediction from financial news articles, IEEE CIVEMSA, pp.60-65, 2017.
- [2] Yu Zheng Zhai, L Hsu, Saman K Halgamuge, Combining News and Technical Indicators in Daily Stock Price Trends Prediction, Advances in Neural Networks, pp.1087-1096, 2007.
- [3] 黒橋禎夫, “JUMAN-KUROHASHI-CHU-MURAWAKI LAB”, <http://nlp.ist.i.kyoto-u.ac.jp/?JUMAN>, (参照2021-01-08)
- [4] 田口雄哉, 田森秀明, 人見雄太, 西鳥羽二郎, 菊田洸, 同義語を考慮した日本語単語分散表現の学習, 情報処理学会第233回自然言語処理研究会, Vol.2017-NL-233, No.17, pp.1-5, 2017.