

深層学習によるニュース記事の重要性と感情指標の 自動推定システムの構築

Development of an estimation system for importance and sentiment measures of news articles using deep learning

野村 理朗^{†§}, 河原 大輔^{*§}, 松尾 正信[§]

Michio Nomura, Daisuke Kawahara, Masanobu Matsuo

[†]京都大学, ^{*}早稲田大学, [§]京都テキストラボ

Kyoto University, Waseda University, Kyoto Text Lab

nomura.michio.8u@kyoto-u.ac.jp

概要

本研究は自然言語処理により、新聞記事に対する読み手の感情と重要性を自動推定するシステムを構築した。はじめにニュース記事の重要性ならびに感情指標に関し、各々2000記事、40000記事からなるオリジナルのデータセットを構築した。続いて深層ニューラルネットワークの先端モデルの一種であるBERTをベースに、記事に対する各心理指標の分類課題を行った結果、感情指標に関わる推定精度は低くとどまったものの、重要性推定の精度は65%とおおよそに実用可能な水準に達することが確認された。

キーワード: 深層学習, 感情解析, 自然言語処理, 新聞記事, 重要性

1. はじめに

感情優先説あるいは感情情報説 (affect-as-information hypothesis) を挙げるまでもなく、感情は潜在的・非意識的に生じたものであっても認知に影響しうる[1, 2]。それは市場において生じたリスクとで同様であり、意思決定へと一定の影響が及ぶ[3]。従来、市場の変動のリスクをヘッジする手法が、金融工学、情報学等の基礎研究分野において、株価等の時系列予測モデル等において開発されてきた[4, 5, 6, 7]。近年は投資家の心理に着眼し、SNSへの投稿内容の感情分析 (sentiment analysis) によりマーケティング、製品レビューに関わる基礎・応用研究が展開していることを受け、ツイッター上の発言内容を金融市場の流動性の予測に活かそうとする試みも複数報じられている[8, 9, 10]。例えばBollenら(2010)は、900万件程度のツイートを収集し、主にはOpenfinderやGPOMS (Google-Profile of Mood States) を基に判定

したツイートの感情極性値等とダウ工業平均株価と時系列変化の関連について検討した結果、特に数日後の株価の下落幅の予測精度が80%程度に及ぶことを報告した。ただし、ここでの予測モデルの検証期間が2008年12月上旬にかぎられているなどの方法論上の限界点に留意したうえで結果を解釈する必要がある。その後、2万3千件程度のツイートを収集し、米国S&P500インデックスとの関連を検討した研究[9]により、ツイートの自然言語処理による感情分析の結果が「流動性(売値買値等から算出される金融指標)」に及ぼす影響はごくわずかにとどまることが報告された。これらの知見に見るように、現時点での感情分析の対象は、ツイートのような比較的短文かつそこに責任や正確性のもめられない個人主体の発言であったり、あるいは一定の予測精度を確立したとしても、予測困難とされるBlack Swanと呼ばれるような事態で生じる膨大な損失をヘッジするような、迅速な対応を可能とする技術には至っていない。さらには英語と日本語との言語構造の相違をふまえ、こうした日本語を対象とした検討を行う必要もある。国内においては、日本語評価極性辞書(名詞編・用言編)[11, 12]を用いたツイートの極性判定が行われている[13]。ここでは経済関連の新聞から抽出した特徴語を基に、経済関連のツイートを発信しているユーザを抽出したうえで、そのツイートの極性判定の結果を素性とすることにより、ユーザを無作為に抽出する条件と比較しての株価の崩落予測精度が向上することが示されている。こうした動向を背景に、本研究は市場のボラタリティに反映される個人や集団の心理の源泉は新聞記事にあるものと仮説し、国内最先端の自然言語処理を駆使することにより、新聞記事のテキストを精確に捉え、過去にない変動をも迅速に予測へと反

表 1 新聞記事データセットにおける記事, 重要性, 重要性推定結果の例

記事	重要性	推定
12日に迫る英総選挙では、ボリス・ジョンソン首相（55）率いる与党・保守党が過半数を獲得し、来年1月末までの欧州連合（EU）離脱を実現できるかどうかが焦点となる	B	✓
*ワシントン=永沢毅*ポンペオ米国務長官は11日、台湾総統選の蔡英文（ツァイ・インウェン）氏の再選を受けて声明を出し「強固な民主主義制度の力強さを再び示した。...	C	✓
バドミントン混合ダブルスで2016年リオデジャネイロ五輪8強の栗原文音（30）＝日本ユニシス＝が今季限りで現役を引退することが21日、分かった。来年4月から同...	E	
アニメ「となりのトトロ」に出てくる「小トトロ」がちょこんとつたキノコ。ブナの木でできた赤や黄色のかさ部分を回すと、アニメで使われた曲「となりのトトロ」や「さ...	D	
みずほ銀行と三井住友銀行は、2020年から一括民営化を始める北海道内7空港を運営する企業連合への3600億円強の協調融資をまとめた。計46の金融機関が参加し、...	C	
松竹の2019年3～11月期の連結営業利益は、前年同期に比べて約4割増の40億円前後だったようだ。映画の話題作が多数公開され、子会社で運営する「MOVIX」な...	C	
*NQNロンドン*6日のロンドン外国為替市場で円相場は小幅続伸した。英国時間16時時点では、前日の同時点と比べて10銭円高・ドル安の1ドル＝108円60～70...	D	✓
函館税関が23日発表した2019年の北海道外国貿易概況（速報）は、輸出総額が3121億900万円と前年比21%減少した。JXTGエネルギーが室蘭市で生産を休止...	D	✓

映させる高精度の自動システムを開発する。その手始めとして、本研究は、新聞記事の重要性や記事に内包される感情価（恐れ、不安、喜び等）を自動的に推定するシステムを構築することを目的とした。

新聞記事データセット

新聞記事の重要性とセンチメントの各々の評定を行った。はじめに重要性としてトレーダー2名（ファンドマネージャ等、投資経験20年以上）が付与した「新聞記事重要性データセット」を構築した。このデータセットは、各記事に対して5段階（A～E; Aが最重要）の重要性を付与したもので、約2,000記事からなるデータセットである。新聞記事重要性データセットに収録された記事の例を表1の第1カラムに、それに対する重要性を第2カラムに示す。

続いてセンチメントとして、クラウドにおいて210名の参加者が100セットの記事（1セット10記事）に対して心理評定を付与した「新聞記事センチメントデータセット」を構築した。このデータセットは、各記事に対する5名の評定者による「恐れ」「不安」「喜び」「経済への期待」等から成る7つの項目についての平均値を付与したもので、約40,000記事からなるデータセットである。新聞記事センチメントデータセットに収録

された記事の例を表 および表 の第1カラムに、それに対する「恐れ」「経済への期待」の値を第2カラムに示す。

2. 方法

新聞記事の重要性推定に関しては、新聞記事1件を入力とし、これに対して重要かどうかの2値分類を行う分類器を構築する。この分類器は文脈言語モデルBERT[14]をベースに作成する。BERTは深層ニューラルネットワークの一種である事前学習付き自己注意機構モデルであり、大規模コーパスを用いた事前学習と対象のタスクにおける fine-tuning の2段階からなる。重要性分類のタスクにおいては、日本語 Wikipedia 全文で事前学習されたモデルを基に、重要性分類タスクにおいて fine-tuning する。この分類は、BERT の[CLS]トークンに対応する出力ベクトルを重要・非重要な2次元ベクトルに変換し、softmax関数を用いることによって、記事が重要・非重要であるスコアを0～1で出力する。訓練は交差エントロピー誤差を用いて行い、テストは重要・非重要なスコアが高い方のラベルを出力する。

新聞記事の感情指標推定に関しては、重要性推定モデルとほぼ同じであるが、出力は各感情指標の実数値

表 2 新聞記事センチメントデータセットにおける記事、「恐れ」の値(正解), 推定値の例

記事	正解	推定
NQNロンドン, 5日のロンドン外国為替市場で英ポンドは対ドルで大幅反発した。英国時間16時の時点で, 前日の同時点と比べて00120ドルのポンド高...	3.4	2.6
バンコク=村松洋兵, タイ政府は電気自動車(EV)関連の新しい投資優遇策を発表した。新規に完成車の生産を始める場合, その事業で生じる法人税の支払いを最長で...	3.2	2.3
在日米大使館はウェブサイトに掲載した3日付の通知で, 日本国内での新型コロナウイルスの感染者数増加を受け, 無期限に海外に滞在するつもりがない限り, 日本に...	3.0	2.4
8日, 国内で新型コロナウイルスへの感染が確認された人が, 514人となり, 一日に感染が確認された人数が初めて500人を超えました。これまで最も多かったのは...	2.4	2.0
トランプ米大統領は7日, 新型コロナウイルス感染拡大を巡り, 世界保健機関(WHO)が「中国中心主義」で, 世界に不適切な提言を行っている」と批判した。...	2.0	2.1
医療費が大きく減っている。4~7月の医療費の総額は前年同期と比べ1兆円規模で減少した。新型コロナウイルス感染症に伴う患者の受診控えや病院の診療体制の縮...	1.6	1.9

表 3 新聞記事センチメントデータセットにおける記事、「経済への期待」の値(正解), 推定値の例

記事	正解	推定
菅義偉首相が10月26日の所信表明演説で「2050年カーボンニュートラル, 脱炭素社会の実現を目指す」と宣言した。昨年12月に就任した欧州連合(EU)の...	3.8	2.8
群馬大学発のベンチャー企業「グッドアイ」(群馬県桐生市)と銅箔(どうはく)系メーカーの「明清産業」(前橋市)が, 付着したウイルスの不活性化を早める銅繊維シートを開...	3.6	2.8
今年で創業100周年を迎えた建設会社のヤマウラ。2つのアルプスに囲まれた長野県駒ケ根市に本社を置き, 県全域で営業する。鉄鋼会社としてスタートしながら, 研究熱心...	3.0	2.5
米連邦準備理事会(FRB)が15日, 再び緊急の金融緩和を決めた。新型コロナウイルスによる経済の混乱を抑えるのが狙いだが, 16日早朝の外国為替市場では円買い圧力...	2.6	2.3
JR四国の経営が厳しさを増している。2021年3月期は12億円の経常赤字を見込み, 黒字化を掲げた中期経営計画との乖離(かいり)を問題視した国土交通省から3月...	2.2	2.4
モデルの蛭原友里が9日, 自身のインスタグラムを更新。自宅で過ごす様子を紹介している。蛭原は「おうち時間をご紹介しますシリーズ, 第2弾がアップされました!」と報告...	1.6	1.9

とする。そのため, BERTの[CLS]トークンに対応する出力ベクトルを1つの実数値に変換し, これを感情指標のスコアとして出力する。訓練は平均二乗誤差を用いて行う。

3. 結果と考察

重要性推定については, 新聞記事重要性データセットを用いて実験を行った。重要・非重要なラベルとしては, 重要性A~Cを重要, D, Eを非重要とした。感情指標推定に関しては, 新聞記事センチメントデータセ

ットを用い, 感情指標としては「恐怖」と「経済への期待」の2つについて推定を行った。

2つのデータともに訓練:開発:テスト=8:1:1に分割し, 訓練データで訓練, テストデータで評価を行った。評価尺度としては, 重要性推定は重要・非重要な精度, 感情指標推定はPearson相関係数およびSpearman順位相関係数を用いた。BERTの事前学習モデルとしては, 「BERT日本語Pretrainedモデル」¹を用い, 入力の記事は形態素解析器Juman++²で単語分割を行った。

重要性推定の精度は65%であった。表1の記事に対

¹ https://nlp.ist.i.kyoto-u.ac.jp/?ku_bert_japanese

² <https://github.com/ku-nlp/jumanpp>

する重要性推定結果を第3カラムに示す。表においてチェック記号が重要と推定された記事を示す。最初の4記事が正答した例で、残りの4記事が誤答した例である。この実験では、重要性A~Cを重要、D、Eを非重要とした2値分類を行ったため、重要性CおよびDの記事について誤答が多かった。重要性推定を改良するためには、どの重要性までを重要とみなすか、また境界の重要性の記事を訓練に使用しないことなどを検討する必要がある。

感情指標推定について、「恐怖」はPearson相関係数が0.213、Spearman順位相関係数が0.186、「経済への期待」はPearson相関係数が0.165、Spearman順位相関係数が0.175であった。表2および表3の記事に対する「恐怖」および「経済への期待」の推定値をそれぞれの表の第3カラムに示す。両方の指標ともに正解の値と推定値は一定の誤差の範囲に収まっていることが多いことがわかる。しかし、推定値は極端な値をとることが少なく、平均値の周辺の値をとることが多かった。この原因としては、テキストのみでは感情指標を推定する手がかりが十分ではないことが考えられる。今後、複数の感情指標を同時に考慮すること、あるいはその精度が従来の機械学習モデルによるものと比較してどの程度にまで高精度であると言えるのかなどを検討する必要がある。

4. まとめと展望

最先端の深層ニューラルネットワークの一種である文脈言語モデルBERTから得られた記事の重要性の推定精度(65%)は、実用に耐えるだろうチャンスレベルを上回るものであった。また金融市場の予測という観点においては、その重要性の高い記事のみを自動抽出しこれを集約するという本研究で行った方法論自体が市場の予測に資するものといえるが、加えて重要性ならびに感情指標を独立変数としたマルチタスクの実施により、市場の動向を高精度に推定するシステムの構築へと結びつくことも想定される。

今後はそうした工学的研究を展開しつつ、それとリンクして感情と重要性の関係性を新たに紐解くことも欠かせない。本研究においてはその目的上、重要性和感情指標の各々について異なる評定者群に付与をもとめたが、今後は参加者内要因計画をデザインし、それらの関連性を解析することが求められる。もとより認知的判断に影響する感情やムード[15, 16]は新聞記事の重要

性判断に対していかなる影響を及ぼすのか。そこにはいかなるプロセスが介在し、それは記事の扱うトピックに応じてどのように異なるのだろうか。今後はこうした問いと実験を積み重ねる中で新たな認知モデルを構築することが期待されよう。

文献

- [1] Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Science*, 11, 393-399.
- [2] Nomura, M., Ohira, H., Haneda, K., Iidaka, T., Sadato, N., Okada, T., & Yonekura, Y. (2004). Functional association of the amygdala and ventral prefrontal cortex during cognitive evaluation of facial expressions primed by masked angry faces: An event related fMRI study. *Neuroimage*, 21, 352-363.
- [3] Guiso L, Sapienza P, & Zingales L. (2018). Time varying risk aversion. *Journal of Financial Economics*, 128, 403-421.
- [4] Choudhry, R. & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39, 315-318.
- [5] Imajo, K., Minami, K., Ito, K., & Nakagawa, K. (2020). Deep portfolio optimization via distributional prediction of residual factors. arXiv preprint, 2012.07245.
- [6] Shah, V. H. (2007). Machine learning techniques for stock prediction. *Foundations of Machine Learning—Spring*, 1, 6-12.
- [7] Ruiz, J. E., Hristidis, V., Castillo, C., Gionis, A. & Jaimes, A. (2012). Correlating Financial Time Series with Micro-blogging Activity, *Proceedings of the fifth ACM international conference on Web search and data mining*, ACM, pp.513-522.
- [8] Bollen, J., Mao, H., & Zeng, X. J. (2011). Twitter mood predicts the stock market, *Journal of Computational Science*, 2, 1-8.
- [9] Zhang, X., Fuhres, H., & Gloor, P (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear”, *Procedia-Social and Behavioral Sciences*, 26, 55-62.
- [10] Guijarro, F., Moya-Clemente, I., & Saleemi, J. (2019). Riquidity risk and investors' mood: linking the financial market liquidity to sentiment analysis through twitter in the S&P500 index. *Sustainability*, 11, 7048-7061.
- [11] 小林のぞみ・乾健太郎・松本裕治・立石健二・福島俊一 (2005). 意見抽出のための評価表現の収集, 自然言語処理, 12, 203-222.
- [12] 東山昌彦・乾健太郎・松本裕治 (2008). 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.
- [13] 佐藤大吾・大原剛三・豊田哲也 (2016). 経済情報に関心をもつSNSユーザの投稿内容に基づく株価騰落予測モデルの構築. SIG-KBS, 1, 1-5.
- [14] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp.4171-4186.
- [15] Schwarz, N, & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.
- [16] Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making. In: Davidson, R J., Scherer, K. R., Goldsmith, H. H., editors. *Handbook of affective sciences*. Oxford: Oxford University Press; pp. 619-642.