

# タグに基づくデータ拡張と因果文抽出を用いた市況コメント生成の試み

## An Attempt to Generate Market Comments via Causal Extraction and Data Expansion Based on Tags

坂地泰紀<sup>1\*</sup> 近藤洋平<sup>2</sup> 和泉潔<sup>1</sup>

長尾慎太郎<sup>2</sup> 加藤惇雄<sup>3</sup>

Hiroki Sakaji<sup>1</sup> Yohei Kondo<sup>1</sup> Izumi Kiyoshi<sup>1</sup>

Shintaro Nagao<sup>1</sup> Atsuo Kato<sup>1</sup>

<sup>1</sup> 東京大学

<sup>1</sup> The University of Tokyo

<sup>2</sup> 大和アセットマネジメント

<sup>2</sup> Daiwa Asset Management

<sup>3</sup> 大和総研

<sup>3</sup> Daiwa Institute of Research

**Abstract:** 資産運用会社や証券会社などの多くの金融機関では、定期的な（月次、週次等）市況コメントを作成し、顧客に提供している。この市況コメントを作成するには、ニュース記事や公的機関発表の統計など、様々な情報源にあたり事実関係を確認しながら、実務者の知識や経験に照らして原稿を作成することになる。一方で、こうした市況コメントの調査・執筆の作業は時間とコストが伴うものである。もし、主要な情報源から、自然言語処理をはじめとした機械的なプロセスによって、信頼できる市況コメントを作成することができれば、それは金融機関における当該業務のコストの削減、あるいは処理の迅速化やカバレッジ拡大による顧客サービス向上に資することになる。そこで、本研究では、まずニュース記事に含まれる原因・結果を含む因果文を抽出することで、市況コメントを自動的に生成することを試みた。

## 1 はじめに

資産運用会社や証券会社などの多くの金融機関では、定期的な（月次、週次等）市況コメントを作成し、顧客に提供している。この市況コメントを作成するには、ニュース記事や公的機関発表の統計など、様々な情報源にあたり事実関係を確認しながら、実務者の知識や経験に照らして原稿を作成することになる。例えば、大和アセットマネジメントが公開している2020年9月の投資環境見通しにおける米株への市況コメントでは、「前回、見通しをやや弱気に引き下げたものの、前回比でS&P500指数は5%弱の上昇と底堅く推移。新型コロナウイルスの感染拡大は継続も、追加経済対策への期待や新型コロナウイルスワクチンへの期待などが株価を支えた。」<sup>1</sup>など、現在の推移と、その原因に

ついて言及しているコメントを作成し、公開している。米国の他にも、ユーロ圏、日本、カナダ、中国、ブラジルなどについてのコメントも記載されている。

一方で、こうした市況コメントの調査・執筆の作業は時間とコストが伴うものである。上記で示した通り、米国だけではなく、カナダや中国などの国について作成する際にも、それぞれの国の株価動向や経済状況などを新聞記事などの情報源から調べ、コメントを生成する必要がある。また、米国や日本などの主要経済国の情報は数多くあるのに対して、ブラジルやインドなどの新興国の情報はそれらに比べて多くはなく、調べるのも難しくなってくる。

もし、主要な情報源から、自然言語処理をはじめとした機械的なプロセスによって、信頼できる市況コメントを作成することができれば、それは金融機関における当該業務のコストの削減、あるいは処理の迅速化やカバレッジ拡大による顧客サービス向上に資することになる。特に、ブラジルやインドなどの新興国の情

\*連絡先：東京大学大学院工学系研究科  
〒113-8656 東京都文京区本郷 7-3-1  
E-mail: sakaji@sys.t.u-tokyo.ac.jp

<sup>1</sup>[https://www.daiwa-am.co.jp/specialreport/outlook/monthly\\_report\\_202009.pdf](https://www.daiwa-am.co.jp/specialreport/outlook/monthly_report_202009.pdf)

報を自動的に集めることができれば、今まで少ない情報を集めていた作業は大幅に削減される。

市況コメントを生成するにあたり、上記の例で示したように市況コメントには、変動の要因を記述することが多い。これは、事前の調査で、作成者が情報源を調べるにあたり、原因とその結果が記述された文を参考に行っている場合が多いということからも、着目すべき点である。そこで、本研究では、まずニュース記事に含まれる原因・結果を含む因果文を抽出することで、市況コメントを自動的に生成することを試みる。

## 2 関連研究

関連研究としては、酒井らの研究 [酒井 19][酒井 20] が挙げられる。酒井らは日経新聞記事を用いて、国内株式市場(日経平均)について月ごとに市況コメントを生成する手法を提案している。彼らは、まず、日経新聞記事から日経平均が大幅に変動したことが記述されている記事のみを抽出し、そこからブーストラップ法に基づいた要因文抽出により、各月の市況コメントを生成している。また、最新の研究発表 [酒井 20] では、doc2vec と因果チェーンによる補完情報を用いることで、生成する市況コメントの質を高めている。それに対して、本研究では、新聞記事に付与されているタグに着目し、日本だけではなく、様々な国の市況コメントを生成できるようなフレームワークを構築している。

また、Murakami et al.[Murakami 17] は日経平均株価の時系列データを入力として、その値動きを用いて市況コメントを生成する手法を提案している。Murakami et al. は、ある短い期間の日経平均株価の値動きから市況コメントを生成しているのに対して、本研究では新聞記事から月ごとの市況コメントを生成しようとしており、タスクが異なる。また、彼らは値動きのみを伝える市況コメントを生成しようとしているのに対して、本研究では因果関係に着目し、値動きの原因を含む形で市況コメントを生成しようとしている点も異なる。

## 3 提案手法

本研究では、以下のステップに従って市況コメントを自動的に生成する。

**Step 1:** タグに基づき各国の市況に関連する新聞記事(関連記事)を抽出する。

**Step 2:** また、Step 1 で抽出した記事を正例にして、機械学習を用いて関連記事を抽出する。

**Step 3:** Step 1 と Step 2 で抽出した関連記事から因果文を機械学習を用いて抽出する。

**Step 4:** 過去の市況コメントを用いて、日次のリターン予測を行う機械学習モデルを学習する。

**Step 5:** Step 3 で抽出した因果文と、Step 4 で学習したモデルを用いて、各文にスコアを付与する。その後、スコアの高い文を市況コメントとして抽出する。 □

ここで、Step 1 で二つのタグを用いる。インド市況に関しては「RR:1122」を、南アフリカ市況に関しては「RR:1163」を用いる。Step 2 ではBERTを用いて関連記事を抽出する。Step 3 の機械学習には、Logistic Regression, Random Forest, SVM, LSTM, LSTM Attention と BERT を試し、最も性能良いものを採用する。Step 4 では、機械学習に Random Forest を用いた。また、Step 5 では、因果文抽出時の尤度と、Step 4 で作成したモデルを用いてリターンの方向性を予測したときの尤度を掛け合わせたものをスコアとして、これに基づいて文をランキングする。

## 4 評価実験

本節では、2つの実験と、市況コメント生成の試みについて述べる。まず一つ目の実験として、因果文の抽出について行った。因果文抽出にはいくつかの手法 [Mirza 14][Dasgupta 18] が提案されているが、どの手法が良いか確認するためにロイターニュース記事(英語)と Semeval2010 Task8 のデータに対して実験を行い、その性能を評価する。ロイターニュース記事は、2,339 文(うち 738 文が因果あり)のデータを学習 7, 検証 1, テスト 2 に分割し利用する。Semeval2010 Task8 は文に含まれる単語間の関係を予測するタスクであり、その関係の一つに原因・結果がある。ここでは、それを正例とし、残りを負例として扱う。Semeval も 10,717 文のデータ(うち 1,331 文が因果あり)を学習 7, 検証 1, テスト 2 に分割し利用する。

二つ目は、日次の市況コメントを入手し、その日の終値に基づいた各国の株式リターンの方向性を予測するものである。日次の市況コメントについては、大和アセットマネジメント様に提供頂いた。市況コメントの中には、因果文であるものもため、全てを使った場合と因果文だけを利用した場合の二通りで実験を行う。因果文の抽出は、前の実験で最もよかったモデルを利用する。これにより、リターンに対する因果文の説明力を確認する。対象とする国は、インドと南アフリカである。インドに関しては、2001 年 9 月 3 日より 2019 年 11 月 29 日までの 4596 件のデータ(うち 4,165 文が因果あり)となっている。南アフリカに関しては、2001 年 9 月 3 日より 2019 年 11 月 29 日までの 4596 件のデータ(うち 3,579 文が因果あり)となっている。デー

表 1: ロイターニュース記事を対象にした因果文抽出結果

	Accuracy	Precision	Recall	F1
Logistic Regression	0.853	0.838	0.825	0.831
Random Forest	0.817	0.829	0.748	0.769
SVM[Mirza 14]	0.813	0.790	0.785	0.787
LSTM[Dasgupta 18]	0.659	0.612	0.608	0.610
LSTM Attention	0.771	0.747	0.768	0.753
BERT	0.877	0.858	0.844	0.866

表 2: Semeval2010 Task8 を対象にした因果文抽出結果

	Accuracy	Precision	Recall	F1
Logistic Regression	0.955	0.899	0.888	0.894
Random Forest	0.945	0.945	0.787	0.844
SVM[Mirza 14]	0.949	0.879	0.880	0.880
LSTM[Dasgupta 18]	0.919	0.801	0.861	0.827
LSTM Attention	0.942	0.873	0.846	0.859
BERT	0.967	0.922	0.922	0.922

タ数が異なるのは、市況コメントがない日が存在するためである。

表 3: 市況コメントを用いた日次リターンの方向予測

	Accuracy	Precision	Recall	F1
インド	0.873	0.876	0.870	0.872
インド (因)	0.893	0.894	0.891	0.892
南ア	0.764	0.765	0.764	0.764
南ア (因)	0.773	0.781	0.773	0.771

## 5 結果と考察

表 1 にロイターニュース記事を対象にした因果文抽出の結果を示し、表 2 に Semeval2010 task8 を対象にした因果文抽出の結果を示す。表 3 に市況コメントを用いたリターンの方向性予測の結果を示す。表 4 と表 5 には、実際に生成した市況コメントの例を示す。

表 1 と表 2 より、因果文抽出に関しては、BERT が最も良い結果となった。そのため、リターンの方向性予測では BERT を用いて実験を行った。表 3 より、因果文に絞ることで、その性能は向上している。このことから、因果文がリターンの方向性予測に対して、説明力が高いことを示唆しており、本提案手法の妥当性を示している。表 4 と表 5 より、インド市場や南アフリカ市場に関連した文を市況コメントとして抽出することができた。今後は、この生成した市況コメントを評価していく。

表 4: インドにおける市況コメント生成例

Indian shares fell on Thursday, dragged by technology stocks, after Infosys Ltd's revenue outlook and plan to return cash to shareholders fell short of expectations, raising concerns at the start of the earnings reporting season.
Indian shares rose on Monday, on track for their second consecutive session of gains, as blue-chips rose after forecast of a timely monsoon raised hopes the central bank would lower interest rates in June.
India's fuel demand rose 6.4 percent in November compared with the same month last year, driven by higher sales of gasoline as discounts and festive season buying boosted passenger vehicle sales.

## 6 まとめ

本研究では、ニュース記事に含まれる原因・結果を含む因果文を抽出することで、市況コメントを自動的に生成することを試みた。BERT により、高精度に因果文を抽出できることを示した。因果文を用いる妥当性として、実験によって、因果文が各国の株式市場リターンの説明力が高いことを示した。リターン予測を行う機械学習と因果文を用いることで、市況コメントの抽出を試みた。今後は、抽出した市況コメントをどのように評価するかを定め、その性能評価を行う。

表 5: 南アフリカにおける市況コメント生成例

---

South Africa's equities market slid on Monday, falling more than 2 percent on the first day of trade in 2016 in the wake of a sharp decline in Asian assets after data showed a contraction in Chinese factory activity.

---

South African shares were led higher by the mining sector on Friday and the market looked set to make further gains in coming sessions after the government slashed fuel prices.

---

South African retailers have struggled to lift earnings as consumers face increasing pressures on disposable income.

---

---

## 参考文献

- [Dasgupta 18] Dasgupta, T., Saha, R., Dey, L., and Naskar, A.: Automatic extraction of causal relations from text using linguistically informed deep neural networks, in *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 306–316 (2018)
- [Mirza 14] Mirza, P. and Tonelli, S.: An Analysis of Causality between Events and its Relation to Temporal Information, in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, pp. 2097–2106 (2014)
- [Murakami 17] Murakami, S., Watanabe, A., Miyazawa, A., Goshima, K., Yanase, T., Takamura, H., and Miyao, Y.: Learning to generate market comments from stock prices, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1374–1384 (2017)
- [酒井 19] 酒井 浩之, 坂地 泰紀, 和泉 潔, 松井 藤五郎, 入江 圭太郎: 関連記事を用いた市況分析コメントの自動生成, 人工知能学会第 22 回金融情報学研究会, pp. 61–66 (2019)
- [酒井 20] 酒井 浩之, 坂地 泰紀, 和泉 潔, 松井 藤五郎, 入江 圭太郎: 学習データ自動生成による市況分析コメント作成のための要因文と補完情報の抽出, 2020 年度人工知能学会全国大会 (2020)