

# 経済レポートを対象とした因果関係に着目する クエリ志向型複数文書要約

高嶺 航<sup>1</sup> 坂地 泰紀<sup>1</sup> 和泉 潔<sup>1</sup> 松島 裕康<sup>1</sup> 島田 尚<sup>1</sup>

<sup>1</sup> 東京大学 大学院工学系研究科

m2018wtakamine@socsim.org

{sakaji, izumi, shimada, matsushima}@sys.t.u-tokyo.ac.jp

## 1 はじめに

近年、投資家に対する投資判断の支援を行う技術の必要性が高まってきている。情報化が進んだ昨今では、投資にあたって参照すべき情報が溢れており、その取舍選択が困難である。

そこで、注目を集めているのが、アナリストレポートである。アナリストレポートとは、金融のプロフェッショナルである証券アナリストが個別企業に対して書くレポートである。IR やニュース、それまでの株価動向、マクロ経済動向などを全て考慮し、アナリスト自身の予想とそれに対する理由が述べられている。そのため、比較的専門性・信頼性が高い情報源である。以上から、アナリストレポートは、投資家達に短時間で効率的に投資に有用な情報を与える情報リソースとして機能している。

しかし、アナリストレポートは、その多くが決算時期に集中して発行される。機関投資家は、極めて限られた期間の中で、投資判断を下さなければならず、集中的に発行される膨大なレポートの全てを熟読し、内容を十分に把握することは難しい。

そこで、本研究では、機関投資家向けの要約システムの構築を目指す。想定するシステムでは、証券アナリストの予想とその理由、根拠の三要素について言及された要約文を提示する。

ここで、想定するシステムを実現する上で、二つの観点が重要である。一つ目として、因果関係に着目して要約を行う必要がある。本システムで提示する予想と理由、理由とその根拠は、結果と原因、つまり因果関係によって繋がっている。これらを適切に抽出するには因果関係を考慮に入れた手法を検討する必要がある。二つ目として、対象のテキストに抽出したい情報が必ずしも書かれているとは限らないことが挙げられる。したがって、探索対象をアナリストレポートに限定せず、証券アナリストが、レポートの執筆時に参

提案システムの出力例

銘柄 企業 K

発行時期 2013年10月29日

予想 利益予想を下方修正

理由 鉱山機械売上が一段と厳しく、回復が遅れる見通し

根拠 燃料炭の国際価格の低迷と足元のインドネシア安の進行などによる影響を受け

根拠ソース 決算短信 (2013年10月28日)

根拠が含まれる文章 <アジア・オセアニア>アジアでは、最大市場のインドネシアにおいて、燃料炭の国際価格の低迷と足元のインドネシア安の進行などによる影響を受け、鉱山機械および一般建設機械の需要が落ち込んだため、売上げは前年同期を大きく下回りました。

図 1: 提案システムの出力例

照するであろう複数の情報リソース (決算短信、新聞記事など) を活用することが重要である。そこで本研究では、アナリストレポート以外の複数種類のテキスト (決算短信、Reuter) を活用し、根拠文の抽出を行った。

提案するシステムでは、図 1 のような要約対象のアナリストレポートとほぼ同時期に発行された決算短信から根拠の情報抽出し、要約文を作成することができた。更に、図 1 の例では、アナリストレポートには具体的には記載されていない根拠情報を決算短信から抽出することができ、本手法の有効性が確認できた。

## 2 研究概要

本研究では、図 2 のようなアーキテクチャを構築した。要約文を出力するまでの過程を図 2 を用いて説明する。

**Step 1** why 型の質問応答のタスクとみなし、証券アナリストの予想とそのアナリストが執筆したレ

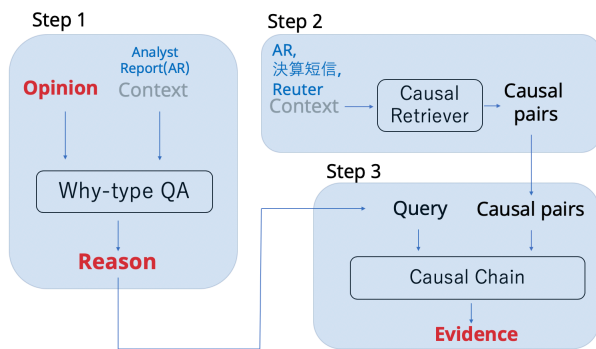


図 2: 全体のアーキテクチャ

ポートの本文の情報から、予想の理由を抽出する。

**Step 2** 予想の理由の根拠文候補を用意するため、アナリストレポート、決算短信、Reuter など複数の金融テキストから大量の因果関係ペアを収集する。

**Step 3** 最後に、抽出した予想の理由をクエリとみなし、Step 2 で獲得した大量の因果関係ペアの中から、適合度の高い因果関係ペアを抽出し、その原因表現を理由の根拠文とする。適合度の判定については、(1) クエリと結果表現との文間類似度と (2) クエリと原因表現との関連度をスコア化し、その値を用いてランキングを行う。

### 3 予想の理由抽出:BERT を利用した Why 型質問応答

質問応答タスクとして、証券アナリストの予想の理由を抽出した。ベースモデルは BERT[1] を採用し、ファインチューニングを行った。今回、取り扱うデータセットは non-factoid 型<sup>1</sup>、特に why 型の質問応答のデータセットである。このようなデータセットは作成コストが高く、データセットのサイズも少量になる場合が多い。

そこで、このような少量のデータセットに対応し、一般公開されており、取得が容易な質問応答のデータセットも活用して、データサイズの拡張の検討を行う。ファインチューニングにおけるデータ拡張の方法として、Garg ら [2] が提唱した 2 段階ファインチューニングを用いた。また、日本語クイズデータセットを基に作成した JaqQA と野村證券株式会社から提供いただいたアナリストレポートを基に作成した OrQA を使用した。

<sup>1</sup>一般に質問応答は、人名や地名、数量などを問う factoid 型質問を対象としたものと、定義・理由・方法等、比較的長い文章表現による記述的な回答が想定される non-factoid 型質問を対象としたものに大別できる。

### 3.1 実験

作成したデータセットを用いて、既存のファインチューニングとの比較を行い、2 段階ファインチューニングの効果を検証する。実験に用いる読解モデルとして、BERT をベースにした 5 つのモデルを採用した。

1. One-shot Fine Tuning in domain : ターゲットのデータセット (OrQA) のみでファインチューニングを行う
2. One-shot Fine Tuning not in domain : ターゲットではないデータセット (JaqQA) のみでファインチューニングを行う。
3. One-shot Mix Fine tuning : ターゲットのデータセット (OrQA) とターゲット外のデータセット (JaqQA) を混合したデータセットを用いて、ファインチューニングを行う。
4. Two-step Fine Tuning( in domain → not in domain) : まず、ターゲットのデータセット (OrQA) でファインチューニングを行い、その次にターゲット外のデータセット (JaqQA) でファインチューニングを行う。
5. (Proposed) Two-step Fine Tuning( not in domain → in domain) : まず、ターゲットではないデータセット (JaqQA) でファインチューニングを行い、その次にターゲットとなるデータセット (OrQA) でファインチューニングを行う。

本タスクのために作成した、JaqQA, OrQA それぞれ計 8835 件、946 件を訓練データ、開発データ、テストデータに分割した。分割比率は 3:1:1 である。用いた日本語 BERT 事前学習モデルは京都大学黒橋・河原研究室が公開しているものを使用する。BERT のファインチューニングでは、一段目、二段目共に Adam を使って学習率  $2e-05$  で最適化する。損失関数は負の対数尤度を用いた。モデルの評価尺度は Exact Match(EM), F1 の 2 つの指標を採用した。EM は出力と正解の完全一致の割合である。F1 は出力された答えの単語列と正解の単語列に対する適合率と再現率から求められる F1 スコアの平均 (F1) を用いた。

### 3.2 実験結果と考察

実験結果を表 1 に示す。表 1 より、2 段階ファインチューニング (JaqQA → OrQA) のモデル (表 1 中の 5) が最も精度が良く、特に、OrQA でファインチューニングのみを行なったモデル (表 1 中の 1) よりも精度が良かったため、2 段階でファインチューニングを行う

表 1: OrQA における実験結果

	data in FPT	data in FT	Dev		Test	
			EM	F1	EM	F1
1	-	OrQA	0.201	0.511	0.196	0.511
2	-	JaQQA	-	-	0.000	0.053
3	-	OrQA & JaQQA	-	-	0.196	0.520
4	OrQA	JaQQA	-	-	0.000	0.057
5	JaQQA	OrQA	<b>0.259</b>	<b>0.545</b>	<b>0.233</b>	<b>0.530</b>

ことの有効性が確認できた。反対に、2段階ファインチューニング (OrQA→JaQQA) のモデル (表 1 中の 4) は、F1 でも、0 に近い値となり、有効に学習できていないことが確認された。そして、ファインチューニング時に、JaQQA および OrQA を同時に使用したモデル (表 1 中の 3) は、2段階ファインチューニング (JaQQA →OrQA) のモデルよりも低いスコアとなった。

したがって、データ拡張を行う際は、単純にデータを混合して使用するのではなく、正しい順序でファインチューニングを行う必要がある。この結果は、Gargら [2] と一致している。

また、2段階ファインチューニングは、専門性の低い業種である、サービス業、卸売業、小売業などにおいて、特に効果があることが確認できている。

## 4 理由の根拠抽出：語義曖昧性を考慮した因果関係抽出

予想の理由の根拠文候補を用意するために複数の金融テキストから因果関係ペアを収集する因果関係抽出器について述べる。テキスト上の因果関係表現の抽出には、坂地ら [3] が提案する手法を採用した。彼らは、因果関係を抽出するうえで、重要な手がかりとなる表現を手がかり表現と定義し、それを利用して、因果関係文の抽出を行う。例えば、「猛暑日が連続したため、飲料水の売上が伸びた。」という文の「ため」が手がかり表現となる。

しかしながら、手がかり表現には、因果関係以外の意味を持つものがある。例えば、「あなたのために、花を買った。」という文中の「ため」は、原因・結果ではなく、目的の意味を表している。このように、手がかり表現を用いる際は、手がかり表現が持つ語義の曖昧性を解消する必要がある。

坂地ら [5] は、このような場合に対応するために、半教師あり学習を用いたフィルタリング手法を適用している。坂地らの手法も分類器の学習を行うために、出現した単語の情報、具体的には、助詞のペア、手がかり表現の直前の形態素の品詞などを含む 6 個の情報を

素性として用いている。しかし、これらは、0 または 1 で素性を表現する離散的な手法であるため、学習データに出現しなかった単語を分類の手がかりとして利用できないという問題がある。

この解決策として、各単語を複数の連続値で構成されるベクトルで表現する方法がある。単語をベクトルで表現する方法は従来から行われてきたが、近年は BERT などの分散表現を素性として用いた手法が提案されている [4]。本研究では、分散表現獲得モデルとして BERT を採用し、獲得した特徴量を基に、手がかり表現の因果関係用法の判別タスクに取り組む。

### 4.1 実験

BERT を利用した、手がかり表現の語義の判定精度を検証するために、3つのデータセットを用意した。それぞれ、語義曖昧性を含む手がかり表現である「で」「から」「によって・により」を手がかりとして、抽出した因果関係文について、因果関係的用法であるか (正例)、そうでないか (負例) のラベルが人手で付与されている。また、各データセットは、訓練データ、開発データ、評価データは、2:1:1 の割合である。また、正例、負例の割合が均等になるように、各データセットで、正例、負例が約 1:1 の割合となっている。

評価尺度として、F1-score を用いた。BERT の pre-training には日本語 Wikipedia すべて (約 1,800 万文) を利用した。

BERT の特徴量の優位性を検証するために、比較手法として、先行研究として挙げた坂地らの素性フィルタリング手法を採用した。他の埋め込み表現の手法として、ELMo, Word2vec(skip-gram) を採用した。手がかり表現の因果関係用法判別器として、SVM, Random Forest, Logistic Regression, MLP を採用した。

### 4.2 実験結果と考察

各手がかり表現の語義曖昧性に関するデータセットを用いた実験結果を表 2 に示す。まず、語義の多義性を考慮しない Word2vec では、どのデータセットに関しても、他のモデルと比較して精度が低い。

先行研究の素性+SVM のモデル, ELMo, BERT の 3つの手法を比較した時、程度は異なるが、どの手法に関しても手がかり表現「から」→「よって」→「で」の順に、精度が落ちている。これにより、語義曖昧性のある手がかり表現の中でも、認識の難易度が異なることが示された。

また、先行研究の素性+SVM のモデルでは「から」のデータセットにおいて高い性能で語義を認識できて

表 2: 3つの手がかり表現の因果関係用法判別タスクにおける比較結果@F1

Embedding	Model	clue:から	clue:よって	clue:で	全体
	<b>Previous Work</b>				
	素性* + SVM [sakaji+ 2011]	0.893	0.784	0.515	0.663
	<b>Word Embedding Based</b>				
Word2vec	SVM	0.663	0.686	0.663	0.508
Word2Vec	Random Forest	0.157	0.364	0.663	0.508
Word2vec	Logistic Regression	0.663	0.686	0.663	0.508
Word2Vec	MLP	0.000	0.686	0.663	0.508
ELMo	SVM	<b>0.968</b>	0.722	0.709	0.751
ELMo	Random Forest	0.950	0.816	0.700	0.865
ELMo	Logistic Regression	0.959	0.864	0.846	0.859
ELMo	MLP	0.959	0.841	0.823	0.882
BERT	SVM	<b>0.968</b>	<b>0.906</b>	0.882	<b>0.916</b>
BERT	Random Forest	0.960	0.853	0.851	0.903
BERT	Logistic Regression	<b>0.968</b>	0.903	0.898	0.910
BERT	MLP	0.952	<b>0.906</b>	<b>0.905</b>	0.902

いるが、「よって」、「で」のデータセットで精度が低くなる。一方、BERT および ELMo を用いたモデルでは、どのデータセットでも 8 割以上の精度で手がかり表現の語義を認識できており、単語の語義を考慮した分散表現の効果が確認できた。そして、BERT + MLP のモデルはどのデータセットにおいても、9 割以上の精度を出しており、特に、最も難易度が高かった「で」のデータセットでは最高精度の成績であった。手がかり表現を用いた因果関係抽出では、「で」の出現割合が最も多いため、運用上は「で」のデータセットで成績が良いモデルを採用するのが適当である。したがって、BERT + MLP は最適なモデルであると考えられる。

## 5 理由の根拠抽出：類似度判定と関連度判定による因果関係接続

前章で獲得した、大量の因果関係ペアを用いて、予想理由の根拠を抽出する。根拠抽出の手法の具体的なステップを確認する。以下の 3 ステップで根拠を抽出することとする。

1. 複数の文書群（今回はアナリストレポート、決算短信、Reuters）から抽出した大量の因果関係ペアの中から、根拠候補となる因果関係ペアの探索を行う。
2. 「証券アナリストの予想の理由」をクエリとみなし、探索した根拠候補の因果関係ペア（原因表現 - 結果表現）について、結果表現とクエリの類似度スコア、原因表現とクエリの関連度スコアを算出する。

3. 類似度スコアと関連度スコアに基づいてランキングを行い、上位 1 位の因果関係ペアの原因表現を「証券アナリストの予想の理由」の根拠とする。

これらの処理を行なった上で、目的とする要約文を獲得する。ただし、これはあくまで要約文を出力するために必要な操作なため、本論文では、その詳細は割愛する。

## 6 考察と結論

本論文では、アナリストレポートの発行時期が集中することが原因で十分に内容を把握することが難しかった機関投資家向けに、投資判断材料に関わる因果関係に着目した要約手法を提案した。具体的には、特に証券アナリストの予想と理由、そしてその根拠を抽出する要約手法となる。本研究では、要約を作成するシステムを、3つのコンポーネントに切り分け、その要素技術について実験を行なった。

本論文では、予想の理由抽出においては、2段階ファインチューニングの適用の有効性が実験的に示せた。また、理由の根拠抽出において、手がかり表現を用いた因果関係抽出手法における語義曖昧性の解消タスクにおいて既存の手法を上回ることができた。今後は、理由の根拠抽出における因果関係接続手法の検証および最終的に出力された要約モデルの検証に取り組むこととする。将来の展望としては、理由の根拠抽出に関する手法の確立および最終的に出力された要約モデルの評価が挙げられる。

## 参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [2] Siddhant Garg, Thuy Vu, and Alessandro Moschitti. Tanda: Transfer and adapt pre-trained transformer models for answer sentence selection. *arXiv preprint arXiv:1911.04118*, 2019.
- [3] 坂地泰紀, 酒井浩之, 増山繁. 決算短信 pdf からの原因・結果表現の抽出. 電子情報通信学会論文誌 D, pp. 811–822, 2015.
- [4] 菅原拓夢, 笹野遼平, 高村大也, 奥村学. 単語の分散表現を用いた語義曖昧性解消. 言語処理学会 第 21 回年次大会 発表論文集, pp. 648–651, 2015.
- [5] 坂地泰紀, 増山繁. 新聞記事からの因果関係を含む文の抽出手法. 電子情報通信学会論文誌 D, Vol. J94-D, No. 8, pp. 1496–1506, 2011.