

表現の置換による業績要因文の同義文・対義文生成

小川 紀寧 酒井 浩之

成蹊大学 理工学部 情報科学科

us212029@cc.seikei.ac.jp, h-sakai@st.seikei.ac.jp

概要

本研究では、決算短信における業績要因文を対象として、文生成にいわゆる大規模言語モデル的手法を使わず、指定の文集合のみを参照して単語・文節の極性を学習し、明示された基準を用いて同義の文や対義の文（意味反転文）を出力する手法を提案した。具体的には、少数の極性提示表現例をもとに、文章中の全文節の極性を自動で算出する。その後、入力文の置換箇所を極性によって定め、置換先の組み合わせ表現を極性や機能語の接続頻度を元に決定し、同義文や対義文の出力を行った。

1. はじめに

現在、生成 AI の発展は目覚ましく、ChatGPT をリリースした OpenAIをはじめ、Google, Amazon, Meta, Apple など多くの企業がモデルを発表し、性能向上に力を入れている。自然言語処理分野でも生成 AI の能力は飛躍的に向上し、ChatGPT に医師国家試験を解かせるテストでは 2023 年には英語で、2024 年には日本語でも最低合格得点率を上回る結果が出た [1]。

しかし、生成 AI による出力はハルシネーションや学習データの偏りによる誤りの可能性が拭えない。ハルシネーションは AI が事実に基づかない情報を生成する現象のことで、これは生成 AI の原理的に排除不可能という主張 [2] も存在する。一方、学習データの偏りはモデルの偏りを生み、出力にモデル特有の本来不必要なバイアスを与えている。実際、GPT によって生成された日本語フェイクニュースのデータⁱを機械学習 (T5 [3]) で自動判別すると高精度・高再現率（両者ともに 97%前後）で判別できた。このデータセットには事実関係を確認しなければ真偽を判断できない文が相当数含まれていたため、機械学

習が文意からニュースの真偽を推定することはできない。つまり GPT の出力には元の文には無い偏りが存在し、機械学習はそれを識別したと考えられる。また AI の出力過程は基本的にブラックボックスである。そのため出力に上記のような問題が発生した場合、解決が難しい可能性が考えられる。

そこで本研究では、指定の文集合のみを参照し、出力プロセスを確認可能な文章生成手法を提案する。参照データの制限は、出力目標に即した文集合のみを参照することで、ノイズとなる情報を記憶させずハルシネーションや不必要なバイアスの抑制を目的としている。また出力プロセスを確認可能とは、出力がどのような経緯で生成されたか説明できることを指し、出力に誤りが存在した際、容易に原因究明できることを目的としている。具体的には決算短信から抽出した業績要因を表す文（以下、業績要因文）12 万文余のデータを参照して、業績要因文に類する文章が与えられた際に同義の文や対義の文を生成した。以下は業績要因文の例である。

またプレミアムヘアケア製品の「ジョン・フリーダ」でも新製品及び改良品を発売し、欧州で順調な伸びを示しましたが、米国では競争激化の影響を受けて伸び悩み、美容サロン向けでは、ほぼ横ばいに推移しました

本手法では業績要因文の「新製品」「改良品」「順調な伸び」といったポジティブ極性表現や「競争激化」「影響を受け」「伸び悩み」といったネガティブ極性表現を推定し、置換先の表現を極性スコアや機能語の接続頻度を元に決定することで文章生成を行う。本手法により説明可能性・確実性の高い文生成を行えるほか、生成の前段階で行う文中出現単語の自動極性分析も、文章内容把握の補佐や文意と一致しない誤り単語検出に寄与すると考える。また、大量の業績要因文に対して同義文、対義文を自動生成することで、そのデータセットは SentenceBERT [4] 等の対照学習のための学習データとしても利用可能であると考えられる。

ⁱ <https://github.com/tanreinama/Japanese-Fakenews-Dataset>

関連研究として、酒井らは、決算短信から業績要因文を抽出する手法を提案しており[5]、本研究で用いる業績要因文は、酒井からの手法に基づいて抽出されたものである。酒井らは業績要因抽出の手がかりとなる極性文節自動抽出も行っており、業績の良し悪しを示す語句（手がかり表現）を自動で抽出している。HuangらやGarciaらは、既存の文を元にした文意が異なる文（偽情報）の自動生成を行っている[6][7]。これらの研究に対して本研究は、極性判定を係り受け関係ではなく各話題内の出現比率によって行う点、文生成ではノイズとなりうるデータを一切参照させないで出力を行い、生成表現選択の過程を確認可能という点で先行研究と異なる。

2. 提案手法

本節では、業績要因文の文脈で同義・対義文を生成する手法の説明を行う。ここで、酒井らの手法によって抽出した業績要因文、約12万文を使用する。また、業績要因文を抽出するための手がかり表現として、業績要因文に含まれる特徴表現167種を用い、それぞれの手がかり表現に対して人手で極性を付与し、極性判定に使用する。以下の例では赤字がポジティブ、青字がネガティブの手がかり表現を表す。

例： **できました** **できませんでした** **回復** **悪化**
堅調 **好調** **不振** **低迷** …

2.1 置換可能な表現の推定

置換可能な表現の推定のために、業績要因文に出現する文節の極性を測り、文節及び機能語の接続頻度を記録する。文節の極性を測る理由は、同義・対義文を作成する際に変更しなければならない表現を推定するためである。ポジティブ・ネガティブ度合がはっきりした表現を入れ替えることで、生成文が元の文に対して同義・対義であることの保証をする。また、機能語（ここでは助詞・助動詞・接続詞）の接続頻度を測る理由は、機能語の方が単語や文節よりバリエーションが少なく、1種類当たりの出現頻度が高いためである。

手法の開発にあたり、以下の3つの仮定を設定した。ここで、文を逆接（ものの、たが、すが、一方）で区切った連文節を「話題」と定義する。

仮定1: 1つの話題には単一の極性(ポジティブ/ニューラル/ネガティブ)が付与できる

仮定2: 話題が極性の異なるものに切り替わる際、逆接を含む文節が必要になる。(逆接: ~ものの、~でしたが、~ですが、~の一方で 等)

仮定3: 手がかり表現自体は、文中に十分存在する→初期状態で十分に**ポジティブ/ネガティブ**の区別をつけられる文節がある

上記の仮定に基づいて、以下の順番で置換可能な表現の推定を行う。

- ① 文を逆接（ものの、たが、すが、一方）で区切った話題（連文節）に分け、手がかり表現を含んでいた連文節に極性を付与する。（手がかり表現が含まれない連文節や、異なる極性の手がかり表現が両方含まれる連文節には付与しない。）

• [またプレミアムヘアケア製品の「ジョン・フリーダ」でも新製品及び改良品を発売し、欧州で**順調**な伸びを示しましたが、] → **ポジティブ**
• [米国では競争激化の**影響**を受けて**伸び悩み**、美容サロン向けでは、ほぼ横ばいに推移しました] → **ネガティブ**

- ② 連文節を文節に分け、各文節がポジティブ・ネガティブな連文節に出現した回数を記録する。

また / プレミアムヘアケア製品の / 「ジョン・フリーダ」でも / 新製品及び / 改良品を / 発売し、 / 欧州で / **順調**な / 伸びを / 示しましたが、 / → **ポジティブ出現数+1**

- ③ これを全業績要因文で行い、文節のポジティブ・ネガティブ連文節出現比率を得る。その後、平均の出現比率と比較して、どちらかの極性に出現が偏っており、偶然そうなる確率が低い文節を拡張手がかり表現として記録する。出現比率および偶然性の基準は付録に示す。顕著に高い/低いポジティブ率と、ポジティブ/ネガティブ手がかり表現偶然性基準の両方を満たす文節を拡張手がかり表現として記録する。入力文が与えられた際、拡張手がかり表現を対応する極性のものに置換することで、同義・対義文を作成する。

2.2 同義文・対義文の生成

文の生成は以下の手順で行う。

- ① 先頭から拡張手がかり表現を探し、前後の文節を確認する。前後の文節が変換後の極性に繋が

るなら、変換しない文節として保存する。繋がらないなら次の文節を確認する。

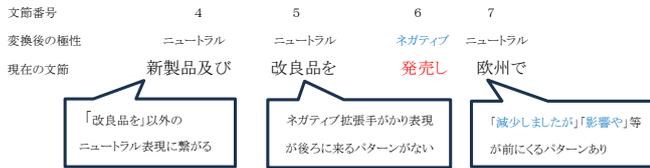


図1 文中の変換候補区間

- ② 保存された文節（図1の例は「新製品及び」と「欧州で」）からそれぞれ、変換後の極性にあった連文節を文節の接続頻度を元に作成する。（図1の例では、新製品及び→ニュートラル→ネガティブ or ニュートラル←ネガティブ←欧州で）
- ③ 各文節番号で、前方枝（図2を参照）が含む文節と後方枝（図3を参照）が含む文節に共通のものが存在しないか確認し、すべて記録する。共通している文節がなかった場合、前方枝と後方枝をどこで繋げるのがもっともらしいか計る接続可能性スコアを、端の文節の出現頻度と機能語接続頻度の積として算出する。各組み合わせの接続可能性スコアを集計し、スコアの上位5つを枝の組み合わせと共に記録する。

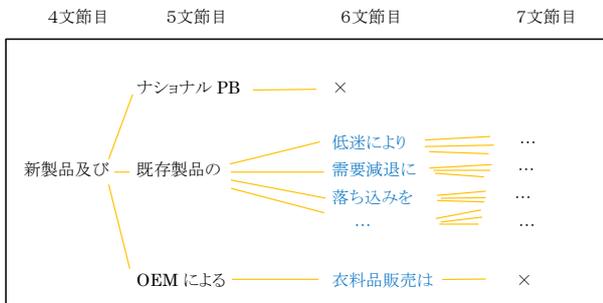


図2 前方文節から生成した連文節（以下、前方枝）集合

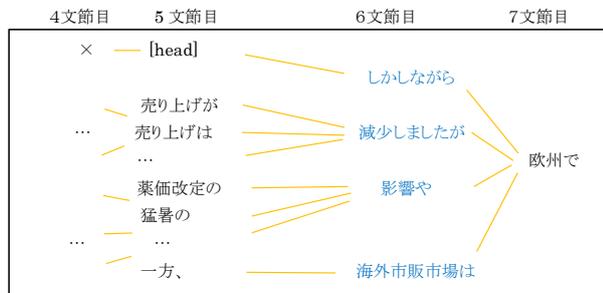


図3 後方文節から生成した連文節（以下、後方枝）集合

- ④ これを1文すべての拡張手がかり表現に対して行う。この時点で各変換箇所につき5パターンの生成連文節が記録されている。文全体ではそれらを組み合わせた5^{変換箇所数}の生成文候補があるので、KenLMⁱⁱによって最終的な選択を行う。
- ⑤ 生成した5^{変換箇所数}通りの出力候補文に対してPerplexityを計り、最もスコアが低いものを最終的な出力文として選択する。

以下に業績要因文の対義文の生成例を示す。

原文

またプレミアムヘアケア製品の「ジョン・フリーダ」でも新製品及び改良品を発売し、欧州で順調な伸びを示しましたが、米国では競争激化の影響を受けて伸び悩み、美容サロン向けでは、ほぼ横ばいに推移しました

対義文

またプレミアムヘアケア製品の「ジョン・フリーダ」でも新製品及び薬価改定の影響や、欧州でやや低調に推移しましたが、米国では順調に売上を順調に増加し、売上高はほぼ順調に推移しました

図4 対義文の生成例

3 評価

上記の手法により生成した12万文余りの対義文の品質を推定する。無作為に選んだ100文に3段階の品質スコア（1:日本語的に問題のない対義文, 2:違和感はあるが文章の体をなしている対義文, 3:対義ではない、もしくは文法的に破綻している文）を人手で評価し、全体の品質を推定する。結果、1:2:3 = 33:51:16 となり、生成文の内8割以上は文法的に問題がなく、3割以上が内容的にも問題ないという結果になった。以下にそれぞれの品質スコアにあたる文を例示する。

ⁱⁱ <https://github.com/kpu/kenlm>

1 包装用フィルムは中国での販売が順調に推移し、前年同期比増収となりました
2 またプレミアムヘアケア製品の「ジョン・フリーダ」でも新製品及び薬価改定の影響や、欧州でやや低調に推移しましたが、米国では順調に売上を順調に増加し、売上高はほぼ順調に推移しました
3 (化学品事業) 主要製品であるクロム塩類、ほう素類ともに、堅調に回復に伴い販売量が急激にことから、当事業の売上高及び営業利益は前年に比べて大幅に下回りました

図5 各品質スコアに割り当てられた生成文の例
紫で問題が発生している箇所を記述

4 考察

評価から、本手法によって文法的に問題がない文を精度よく生成することができた。ここで生成した同義文・対義文が機械学習で判別できないかどうかを確かめるために、T5による判別器での検証を行った。評価方法は交差検定を用いた。まず前項で作成した文生成プログラムで、業績要因文12万文余りに対応する同数の出力文を生成する。元データと生成文を分割し、約3万文を1セットとして8セットのデータ群を用意した。そのうち7つを訓練データとしてT5が生成文を見分けられるように学習を行い、残った1セットのテストデータを判別する。なお、対義文ではない生成文は評価から除外した。本評価では、生成文をT5が判別できないことが望ましい検証となる。

表1 T5による判定結果

セット	生成文と判定した文数	原文と判定した文数	再現率	精度
1	14838	15060	0.983	0.989
2	14838	15005	0.984	0.985
3	14840	14999	0.983	0.984
4	14852	14961	0.983	0.982
5	14863	15076	0.983	0.989
6	14834	15017	0.983	0.986
7	14862	15046	0.982	0.988
8	14873	15003	0.984	0.985

結論としては、再現率、精度ともに98%を超える高水準で判別されてしまった。なぜT5で判別され

たかについて考察する。まず判別された出力例を見ると、複数文節を組み合わせたときの整合性に問題があるように見受けられる。例えば図4には「プレミアムヘアケア製品」の話題で「薬価改定」が出てきている点、並列助詞である「や」の前後が同じ種類の句ではない点、「順調に/売りを/順調に」といった、2種類の連文節(～順調に売上を～、～売上を順調に～)が中央の文節で混ざっている点になる。

なお、T5で生成文と判別されなかった文も2045文(約1.7%)存在する。T5で判別されなかった生成文は①置換箇所が一つの短文②業績結果だけを述べる文③置換箇所が連続していない文(図5の1など)で占められていた。つまり2.2節で作成する枝が短いほど破綻した文が発生しにくいことが伺える。

ポジティブな見方をすれば、機能語の接続関係を測ることで大規模言語モデルを使わなくても単文節間の文法破綻は防げることが伺えた。よって出現頻度の低い名詞の接続関係把握や連文節の文法的妥当性が計れるならば、大規模言語モデルなしに破綻しない文章を作成できると予想する。前述した「名詞の接続関係把握」は、属するカテゴリが違う単語を一つの話題にいれないことを目標にしている。方針としては3つ考えられ、1つは単語のクラスタリングを行い、単語や文節をグループ分けすること、2つ目はどんな話題にもある程度の汎用的な語句のみを生成に使うこと、3つ目はカテゴリを持ちうる名詞(固有名詞・一般名詞)を変換対象から外すことである。

5 むすび

本研究では、文生成にいわゆる大規模言語モデルを使わず、指定の文集合のみを参照して文節の極性を推定し、明示された基準を用いて同義文や対義文を出力する手法を提案した。具体的には、少数の極性手がかり表現をもとに、文章中の全文節の極性を自動で算出する。その後、入力文の置換箇所を極性によって定め、置換先の組み合わせ表現を極性や機能語の接続頻度を元に決定し、同義文や対義文の出力を行う。評価実験では、人手による評価では十分な精度を達成することができた。ただし、T5による自動判別では高精度で判別されてしまい、生成された文には不自然な点があることが伺える。今後は、名詞の接続関係を考慮し、機械学習でも判別できない文の生成を目指す。

参考文献

- [1] Yudai Tanaka, Takuto Nakata, Ko Aiga, Takahide Etani, Ryota Muramatsu, Shun Katagiri, Hiroyuki Kawai, Fumiya Higashino, Masahiro Enomoto, Masao Noda, Mitsuhiro Kometani, Masayuki Takamura, Takashi Yoneda, Hiroaki Kakizaki, Akihiro Nomura, “Performance of Generative Pretrained Transformer on the National Medical Licensing Examination in Japan”, *PLOS Digit Health*, vol.3, no. 1, 2024.
- [2] Sourav Banerjee, Ayushi Agarwal, Saloni Singla, “LLMs Will Always Hallucinate, and We Need to Live With This”, arXiv:2409.05746v1, 2024.
- [3] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, *Journal of Machine Learning Research* 21, pp.1-67, 2020.
- [4] Nils Reimers, Iryna Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3982–3992, 2019.
- [5] 酒井浩之, 松下 和暉, 北島 良三, 学習データの自動生成による決算短信からの業績要因文の抽出, *日本知能情報ファジィ学会誌*, vol. 31, no. 2, pp. 653-661, 2019.
- [6] Kung-Hsiang Huang, Kathleen McKeown, Preslav Nakov, Yejin Choi, and Heng Ji, “Faking Fake News for Real Fake News Detection: Propaganda-Loaded Training Data Generation”, In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pp. 14571–14589, 2023.
- [7] Gabriel Lino Garcia, Pedro Henrique Paiola, Danilo Samuel Jodas, Luis Afonso Sugi, and João Paulo Papa, “Text Summarization and Temporal Learning Models Applied to Portuguese Fake News Detection in a Novel Brazilian Corpus Dataset”, In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, vol.1, pp. 86–96, 2024.

付録

A. ポジティブ・ネガティブ手がかり表現基準

$$(1) \text{文節ポジティブ率 } p_i = \frac{\text{ポジティブ連文節出現数 } ps_i}{\text{ポジティブ連文節出現数 } ps_i + \text{ネガティブ文節出現数 } ng_i}$$

$$(2) \text{平均ポジティブ率 } p_m = \frac{\sum ps_i}{\sum (ps_i + ng_i)}$$

$$(3) \text{顕著に高いポジティブ率 } p_{\text{high}} \geq 1 - \frac{(1 - \text{平均ポジティブ率 } p_m)}{2}$$

$$(4) \text{顕著に低いポジティブ率 } p_{\text{low}} \leq \frac{\text{平均ポジティブ率 } p_m}{2}$$

$$(5) \text{二項分布 } f(x, n, p) = {}_n C_r p^x (1-p)^{n-x}$$

$$(6) \text{二項分布下側累積確率 } P(x, n, p) = \sum_{t=0}^x f(t, n, p)$$

$$(7) \text{二項分布上側累積確率 } Q(x, n, p) = \sum_{t=x}^n f(t, n, p)$$

$$(8) \text{ポジティブ手がかり表現偶然性基準: } 0.1 > Q(ps_i, ps_i + ng_i, p_m) - \frac{f(ps_i, ps_i + ng_i, p_m)}{2}$$

$$(9) \text{ネガティブ手がかり表現偶然性基準: } 0.1 > P(ps_i, ps_i + ng_i, p_m) - \frac{f(ps_i, ps_i + ng_i, p_m)}{2}$$

文節 i が (3) と (8) を満たすならポジティブ手がかり表現, (4) と (9) を満たすならネガティブ手がかり表現と見なす