

# 決算短信からの業績要因文の抽出に向けた 業績発表記事からの訓練データの生成

大村 和正<sup>1</sup> 白井 穂乃<sup>2</sup> 石原 祥太郎<sup>2</sup> 澤 紀彦<sup>2</sup>

<sup>1</sup> 京都大学大学院情報学研究科 <sup>2</sup> 株式会社日本経済新聞社

omura@nlp.ist.i.kyoto-u.ac.jp

{hono.shirai,shotaro.ishihara,norihiko.sawa}@nex.nikkei.com

## 概要

決算短信からの業績要因文の抽出には一定の需要があり、業績に関わる要因が記述された文を高精度に抽出することができれば投資支援として非常に有用である。本稿では、業績発表記事から単純なルールで訓練データを生成し、決算短信からの業績要因文の抽出に向けた深層学習モデルを構築する。提案手法はデータ構築が容易であるという特徴を持ち、再現率の高い抽出モデルが構築されることを示す。

## 1 はじめに

ウェブの普及に伴って日々膨大な情報が電子データとして蓄積されるようになり、各々の需要に合わせて必要な情報を抽出する技術はますます重要になっている [1, 2]。このような場面の一例として、決算短信からの業績要因文の抽出 [3, 4] がある。

決算短信とは、上場企業が決算発表を行う際に開示する、当期の業績等をまとめた書類である。投資家は決算短信に記載された情報から企業の動向を推測し投資行動の参考にするため、決算短信は極めて重要な情報源である。一方で、決算発表が義務付けられている上場企業数は約 4,000 企業にのぼり、その多くは年に 4 回決算発表を行うため、これらの決算短信を把握するには多大な労力を要する。企業の動向の判断材料となる、業績に関わる要因が記述された文を自動抽出することができれば、負荷軽減および即時性の観点から投資家への支援として非常に有用であると考えられる。

本研究では、中山らの定義 [5] を参考に、決算短信に含まれる文を以下のように分類定義する。

**業績文** 当期の業績変化のみを述べた文

**要因文** 文内で当期の業績変化と明示的に関連付けられていないが、要因であると判断される文

表 1 ある企業の決算短信と業績発表記事の例 (決算短信は定性的情報から、業績発表記事は冒頭 1 段落を抜粋)。

決算短信	... また一方で、巣ごもり消費と呼ばれる新しい消費行動が拡大し、感染対策で消費者の清潔志向も高まるなど、新しい需要が創出されたことで、それらに対応した商品は売上が大幅に増加しました。衣料品の販売に影響を及ぼす天候は、9月中旬まで残暑が続いたものの、9月下旬には全国各地で最低気温が 20℃を下回るなど秋らしい気温となりました。...
業績発表記事	<企業名>が 28 日発表した 2020 年 9~11 月期の連結決算は、純利益が前年同期比 3.6 倍の 110 億円だった。9~11 月期として過去最高となる。気温低下で秋冬物の販売が好調だったほか、値下げの抑制で粗利が改善した。新型コロナウイルス禍を受けた巣ごもり消費で、部屋着やインテリア商品などの需要も堅調だった。

**業績要因文** 文内で当期の業績変化とその要因が述べられた文

**その他** 注記など、上記のいずれにも該当しない文。

この中で、業績要因文および要因文の抽出を目的とする。例えば、表 1 上段の一文目は「売上が大幅に増加」という業績変化とその要因が文内で関連付けて述べられているため、業績要因文である。

本稿では、決算短信からの業績要因文の抽出に向けて、業績発表記事から深層学習モデルの訓練データを生成する手法を提案する。業績発表記事とは、決算発表後に第三者 (記者) によって作成される決算短信の要約記事である。表 1 下段のように、業績発表記事には当期の業績と記者が重要だと判断した業績要因が簡潔に記述されており、要因の分類基準の学習に適している。また、重要文を学習できるため、業績発表記事を模した決算サマリー<sup>1)</sup>の自動生成への応用を視野に入れることができる。

提案手法では、業績発表記事から定性的な文を単純なルールで判別し、これを要因文として学習する。データ生成を工夫することで、再現率の高い抽出モデルが構築されることを実験的に示す。

1) <https://www.nikkei.com/promotion/collaboration/qreports-ai/>

## 2 関連研究

酒井らは、手がかり表現と企業キーワードによる業績要因文の抽出手法を提案した [6, 4]. 手がかり表現は「好調でした」のような、業績要因となる状態や変化を表す用言的な表現と定義され、業績発表記事または決算短信から以下の手順で獲得する。

1. 少数の手がかり表現を手で与え、その手がかり表現に係る文節を取得する。
2. 取得した文節のうち、様々な手がかり表現に一樣に係るものを抽出する (共通頻出表現と呼ぶ)。
3. 共通頻出表現に係る文節を取得し、その中で様々な共通頻出表現が一樣に係るものを新たな手がかり表現として獲得する。
4. 上記の操作を一定回数繰り返す。

例えば、2で係り先の一様性を測る時は、共通頻出表現の候補  $f$  が手がかり表現  $c$  に係る確率  $P(f, c)$  にもとづくエントロピー  $H(f)$  を用いる。これらは次のように計算される。

$$P(f, c) = \frac{\#(f, c)}{\sum_{c' \in C} \#(f, c')}$$
$$H(f) = \sum_{c' \in C} -P(f, c') \log_2 P(f, c')$$

ただし、 $\#(f, c)$  は  $f$  が  $c$  に係る回数、 $C$  は手がかり表現の集合であり、係り先が一樣であるほど  $H(f)$  が高くなる。3でも同様に係り元の一様性を測る。

企業キーワードは商品名や部門名など、各企業にとって重要な名詞句を指す。ある企業  $t$  の決算短信集合  $S(t)$  における名詞句  $n$  の重要度  $W(n, S(t))$  を次のように定義し、これをもとに自動獲得する。

$$W(n, S(t)) = tf(n, S(t)) \times \log_2 \frac{N}{df(n)} \times H(n, S(t))$$

ただし、 $tf(n, S(t))$  は  $S(t)$  における  $n$  の頻度、 $df(n)$  は  $n$  を含む決算短信がある企業数、 $N$  は収集した決算短信の企業数、 $H(n, S(t))$  は  $S(t)$  における  $n$  の出現確率にもとづくエントロピーである。  $S(t)$  を1つの文書とみなした時の tf-idf が大きく、 $S(t)$  の中で満遍なく現れる単語ほど重要度が高くなる。

業績要因文の抽出時は、手がかり表現を含み、その係り元に企業キーワードを含む文を抽出する。マッチングベースの手法であり、再現率が低いという問題が指摘されている [7].

中山らは、業績要因文を定義し、酒井らの手法を発展させて文間の結束性を考慮する業績要因文の抽

表2 評価データの統計.

	開発	テスト
業績要因文または要因文である	48	179
// でない	144	999
総文数	192	1,178

出手法を提案した [5]. 文間の談話関係や語の重複を手がかりに抽出結果を更新することで、精度を改善している。

酒井らは、前述の手法を深層学習モデルの訓練データの自動生成に応用することで、より再現率の高い業績要因文の抽出手法を提案した [7]. 決算短信から確信度の高い業績要因文の候補を抽出し、これを訓練データとして利用するため、適合率を落とさず再現率を向上させている。本稿では、業績発表記事から分類基準を学習することを目的としており、この点で若干異なる。

## 3 提案手法

提案手法では、業績発表記事から定性的な文を「要因文」、定量的な文を「業績文」として抽出することで訓練データを自動生成する。定性的であるか否かは文中に数字を含むか否かという単純なルールで判別する。具体的には以下のような手順で訓練データを自動生成する。

1. 業績発表記事を文単位に分割し、各文に対して、数字を含む文を「業績文」、含まない文を「要因文」と自動でラベル付けする。
2. 冒頭1段落に含まれる業績文と要因文を適当な接続語<sup>2)</sup>でつなぐことで疑似的な業績要因文を生成する。

2について、表1下段の例から「気温低下で秋冬物の販売が好調だったほか、値下げの抑制で粗利が改善したことにより<企業名>が28日発表した2020年9~11月期の連結決算は、純利益が前年同期比3.6倍の110億円だった。」といった文が生成される。

## 4 実験

### 4.1 データ構築

**業績発表記事の取得** 手がかり表現の獲得および深層学習モデルの訓練に利用する業績発表記事は、

2) 本研究では、「ことで」「ことにより」「こともあり」の3つを用いた。

表3 訓練データの統計.

業績文 (定量的な文)	26,197
要因文 (定性的な文)	17,696
擬似的な業績要因文	8,824

株式会社日本経済新聞社が提供するニュース配信サービス「日経電子版」<sup>3)</sup>から取得した。具体的には、2016年1月1日から2020年12月31日までの5年間を対象とし、メタデータのトピック情報に「企業決算」ラベルが付与されているものを取得した。取得された記事の総数は3,322件、総文数は43,893文、1文あたりの平均文字数は39.8文字であった。

**決算短信の取得** 企業キーワードの抽出および各手法の評価に利用する決算短信は、公開されている決算短信から抽出したテキストを保存している内製のデータベースから取得した。業績発表記事の取得時と同様に2016年1月1日から2020年12月31日までの5年間を対象とした。取得されたデータのうち、タイトルに「決算短信」の文字列を含まないものや本文が6,000文字未満<sup>4)</sup>のものなどを除き、最終的に6,067件の決算短信を獲得した。

獲得された決算短信のテキストデータには表を無理矢理テキスト化した乱雑な文字列を含むため、酒井らの手法[4]を参考に次のようなクレンジングを行った。

1. テキストを句点「。」で分割する。
2. 分割された文字列のうち、350字以下のものを文として抽出する。
3. 350字より多いものは全角スペースまたは下駄(■)で分割し、分割された末尾の文字列が350字以下であれば、それを文として抽出する。

**評価データの作成** 獲得した決算短信から開発用に5件、テスト用に20件を無作為に抽出し、<sup>5)</sup>各文に対して「業績要因文または要因文である」か否かのアノテーションを行った。この評価は著者らが人手で行い、3人以上の合意があったものを正例とした。評価結果を表2に示す。

## 4.2 実装の詳細

決算短信には次期の業績予想や注記など、当期の業績に関する定性的情報と無関係な文が数多く含ま

れる。このような文を事前に除外するために、各文に次のようなルールを適用した。

1. 文末の節が過去形でない文は「業績要因文または要因文でない」
2. 「努めました」「はかりました」「まいりました」のいずれかで終わる文は「業績要因文または要因文でない」
3. 「店舗数」の文字列を含む文は「業績要因文または要因文である」

これらのルールでラベル付けできなかった文に対し、ルールのみ・ベースライン・提案手法の3つの設定でラベルを予測した。精度は自動でラベル付けされた文を含めて算出した。

**ルールのみ** 上記のルールを満たさない文は全て「業績要因文または要因文である」とする。ルールそのものの影響を定量的に測るために行う。

**ベースライン** 本実験では、手がかり表現と企業キーワードによる業績要因文の抽出手法[6, 4]をベースラインとした。2節の順に従って、業績発表記事3,322件から手がかり表現を、決算短信6,067件から企業キーワードを獲得した。文節および係り受けの解析にはKNP[8]<sup>6)</sup>を利用した。使用した手がかり表現の数は202個であった。

手がかり表現は「好調だった」のように付属語を含みうるため、常体の文章(業績発表記事)から獲得した手がかり表現を敬体の文章(決算短信)とのパターンマッチにそのまま用いるのは適切でないと考えられる。そのため、JUMAN辞書の代表表記[9]を用いて手がかり表現を正規化した際の精度も検証した。

**提案手法** 3節の手法を適用し、深層学習モデルの訓練データを構築した。文分割はpython-textformattingライブラリ<sup>7)</sup>を利用した。訓練データの統計を表3に示す。

素朴にこの訓練データを用いると、数字を手がかりとするモデルが構築されてしまうと考えられる。これに対処するため、業績文に含まれる数字を空文字列に置換して[10]訓練した時の精度も検証した。

業績要因文の分類モデルとしてBERT[11]を用いた。BERTの事前学習モデルは、日本語Wikipediaで事前学習したNICT BERT 日本語 Pre-trained モデル<sup>8)</sup>を利用した。fine-tuningはDevlinらが提案した

3) <https://www.nikkei.com/>

4) 本文が比較的短いものは当期の業績に関する定性的情報を含まないものが大半を占めていたため、この条件を設定した。

5) 同一企業の決算短信が複数含まれていないことを確認した。

6) <https://github.com/ku-nlp/KNP>

7) <https://github.com/ku-nlp/python-textformatting>

8) <https://alaginrc.nict.go.jp/nict-bert/index.html>

表4 テストデータに対する実験結果. 提案手法については, 異なる3つのシード値で fine-tuning した結果の平均と標準偏差を記載している.

手法		正解率	適合率	再現率	F 値
ルールのみ		73.2	35.4	92.7	51.2
ベースライン [6, 4]	+手がかり表現の正規化	88.8	67.7	46.9	55.4
		90.5	<b>70.7</b>	62.0	66.1
提案手法		89.4 ± 0.7	68.7 ± 1.6	55.9 ± 6.0	61.5 ± 4.2
	+数字除去	89.4 ± 1.0	69.8 ± 4.3	52.9 ± 3.0	60.2 ± 3.4
	+擬似的な業績要因文	90.2 ± 0.6	64.4 ± 1.8	79.3 ± 1.2	71.1 ± 1.6
	+数字除去+擬似的な業績要因文	<b>91.3 ± 0.8</b>	68.4 ± 2.6	<b>80.1 ± 2.5</b>	<b>73.8 ± 2.4</b>

表5 開発データに対する BERT モデルの誤分類例. 「業績要因文または要因文である」ことを「正」, そうでないことを「負」と表現した. また, 著者らが要因と判断した部分を太字で表示している.

予測	正解	文
正	負	当第3四半期連結累計期間の日本経済は, 新型コロナウイルス感染症拡大の影響を大きく受ける結果となりました.
正	負	このような環境のもと, 当社グループは, 従業員のマスク着用, レジ及び服薬指導カウンターへの飛沫感染予防フィルターの設置, 店舗出入口への消毒用アルコールの常設など感染予防対策を徹底し, 地域社会の人々の生活を支える社会インフラとしての役割を担うべく営業活動を継続いたしました.
負	正	<b>4月下旬から約2週間は国内実店舗の全てが休業し</b> , その後緊急事態宣言の解除に伴い順次営業を再開いたしました.
負	正	以上のような諸施策に取り組みましたが, 前年同四半期に比べ売上収益は43.9%減の493億91百万円, 第2, 第3四半期連結会計期間は黒字を確保できましたが, <b>全館休業の影響が大きく</b> , 営業損失は15億79百万円(前第3四半期連結累計期間は営業利益101億30百万円)となりました.

単文分類の設定 [11] で行い, 要因文および擬似的な業績要因文を正例, 業績文を負例とする2値分類を学習した. ハイパーパラメータの詳細は付録 A.2 に記載した. 開発データに対する精度をエポックごとに調べ, F 値が最も高かった時のパラメータでテストデータに対する精度を検証した.

### 4.3 実験結果

テストデータに対する実験結果を表4に示す. ベースライン手法は, 正規化した手がかり表現を用いることで適合率・再現率ともに向上している. これは, 文体のミスマッチにより抽出できない業績要因文が数多く存在する一方で, 業績発表記事の情報が決算短信からの業績要因文の抽出にある程度有用であることを示している.

提案手法は, 擬似的な業績要因文を訓練データに加えることで再現率が大きく向上している. 業績発表記事では業績とその要因が別々の文に分けて記述されることが多いため, 決算短信に特徴的な「文内で要因を述べてから業績を述べる」パターンが擬似的な業績要因文によって学習されたと考えられる.

また, 数字を空文字に置換するデバインシングの有無に着目すると, これによって適合率が多少向上

していることが分かる. 数字以外に着目させることで, 定量的な文と定性的な文がより判別できるようになったと考えられる.

### 4.4 定性的分析

開発データの分類結果をもとに提案手法のモデルが誤分類する文の傾向を分析した. 誤分類の例を表5に示す.

- 国家的もしくは世界的な経済状況を述べた文
- 定性的ではあるが, 取り組みとして弱いために要因文と評価されなかった文
- 定量的な表現を多く含む文

が誤分類される傾向にあった.

## 5 おわりに

本稿では, 業績発表記事から訓練データを生成し, 決算短信から業績に関わる要因が記述された文を抽出する深層学習モデルの構築に取り組んだ. 今後の課題としては, データ構築手法の改善や評価データの拡張, 決算短信を訓練データに利用した場合との比較などが挙げられる. また, 抽出精度の改善と並行して, 極性の予測や決算短信を要約した決算記事の生成手法も検討したい.

---

## 参考文献

- [1] Susumu Akamine, Daisuke Kawahara, Yoshikiyo Kato, Tetsuji Nakagawa, Kentaro Inui, Sadao Kurohashi, and Yutaka Kidawara. WISDOM: A Web Information Credibility Analysis Systematic. In **Proceedings of the ACL-IJCNLP 2009 Software Demonstrations**, pp. 1–4, Suntec, Singapore, August 2009. Association for Computational Linguistics.
- [2] Junta Mizuno, Masahiro Tanaka, Kiyonori Ohtake, Jong-Hoon Oh, Julien Kloetzer, Chikara Hashimoto, and Kentaro Torisawa. WISDOM X, DISAANA and D-SUMM: Large-scale NLP Systems for Analyzing Textual Big Data. In **Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations**, pp. 263–267, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.
- [3] 西沢裕子, 酒井浩之. 企業の決算短信 pdf からの業績要因の自動抽出. 電子情報通信学会技術研究報告 = IEICE technical report : 信学技報, 2013.
- [4] 酒井浩之, 西沢裕子, 松並祥吾, 坂地泰紀. 企業の決算短信 pdf からの業績要因の抽出. 人工知能学会論文誌, Vol. 30, No. 1, pp. 172–182, 2015.
- [5] 中山祐輝, 津々見誠, 村上浩司. 文間の結束性に基づく決算短信における業績要因文の抽出. 言語処理学会 第 25 回年次大会, 2019.
- [6] Hiroyuki Sakai and Shigeru Masuyama. Cause Information Extraction from Financial Articles Concerning Business Performance. **IEICE Transactions on Information and Systems**, No. 4, pp. 959–968, 2008.
- [7] 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. 知能と情報, Vol. 31, No. 2, pp. 653–661, 2019.
- [8] Sadao Kurohashi and Makoto Nagao. A Syntactic Analysis Method of Long Japanese Sentences Based on the Detection of Conjunctive Structures. **Computational Linguistics**, Vol. 20, No. 4, pp. 507–534, 1994.
- [9] 黒橋慎夫. 言語のセマンティックス. 人工知能, No. 6, pp. 718–723, 2006.
- [10] Brendan Kennedy, Xisen Jin, Aida Mostafazadeh Davani, Morteza Dehghani, and Xiang Ren. Contextualizing Hate Speech Classifiers with Post-hoc Explanation. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, July 2020.
- [11] 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: Human Language Technologies, Volume 1 (Long and Short Papers)**, 2019.

## A 実装の詳細の補足

### A.1 ベースライン

手がかり表現のシードは酒井らの手法 [6, 4] にならない, 「が好調」と「が不振」の2つを利用した. また, 手がかり表現は用言の節, 共通頻出表現は体言の節に限定し, 数字を含むものは除外した. 共通頻出表現および手がかり表現の獲得に関するエントロピーの閾値のパラメータ  $\alpha, \beta$  は {0.5, 0.625, 0.75} の3通りを試し, 最終的に 0.625 に設定した. また, 2節で述べたステップ 1~3 は 3 回繰り返した. 獲得された手がかり表現の例を表 6 に示す.

表 6 手がかり表現の獲得例. 括弧内は Juman 辞書の代表表記 [9] を表す.

改善した (改善/かいぜん)
伸ばした (伸ばす/のばす)
堅調だった (堅調だ/けんちょうだ)
拡大し (拡大/かくだい)
増えている (増える/ふえる)
苦戦した (苦戦/くせん)
伸び悩んだ (伸び悩む/のびなやむ)
悪化した (悪化/あつか)
減少し (減少/げんしょう)
低迷している (低迷/ていめい)

### A.2 提案手法

ハイパーパラメータの詳細は表 7 のとおりである. 業績発表記事に対して過学習すると, 決算短信に対する抽出精度が悪化すると考えられるため, 学習率を一般的な値より低く設定している.

表 7 ハイパーパラメータの詳細.

パラメータ名	パラメータ値
バッチサイズ	32
エポック数	3
学習率	1e-7
最大トークン長	128
Optimizer	AdamW
Adam's betas params	(0.9, 0.999)
Adam's epsilon	1e - 6
重み減衰	1e-2
Scheduler	Linear decay (with no warmup)
シード値	{0, 1, 2}