

業績要因を用いた決算短信のタイトル自動生成

今井 康太, 高野 海斗, 酒井 浩之
成蹊大学

us162018@cc.seikei.ac.jp, dd186201@cc.seikei.ac.jp, h-sakai@st.seikei.ac.jp

1. はじめに

近年, 証券市場における個人投資家が増加しており, その影響として個人投資家への支援を行うシステムの需要が高まっている. 個人投資家は複数ある上場企業の中からどこに投資をするかを判断する必要がある. どの企業に投資をするかを判断する資料として, 企業が業績発表を行った後に公開される決算短信が存在する. 決算短信は上場企業の決算情報が最も早く開示される資料であり, 投資判断において重要な資料である. しかし, 公開される決算短信は数が多く, 人手で全ての業績結果や業績要因を把握することは困難である.

支援システムとして決算短信から自動で記事とそのタイトルを生成するシステムが公開されている[1].しかし, タイトルには業績結果のみが記載されており, なぜその結果になったのかが把握することができない.

そこで本研究では業績要因を含む文に注目した.(以降, 業績要因を含む文を「業績要因文」と定義する.)業績要因文とは, 例えば「太陽電池の販売が順調に拡大した」のような文であり, 何が企業の業績結果に影響を及ぼしたかを読み取ることができる. 経済新聞における業績発表記事のような, 実際に人手によって作成された記事のタイトルには業績要因が含まれていることが多く, 決算短信のタイトルにおいても業績要因を把握できる業績要因文は重要だと考えられる. 本研究では決算短信から業績結果と業績要因文を抽出し, それらを利用して, 例えば「フューチャー, 純利益 123.8%増, ITコンサルティング事業が好調」のような決算短信のタイトルを自動生成する手法を提案する.

関連研究として, 山本らは一般的な新聞記事を新幹線車内や街頭の電光掲示板で流れるニュースのような簡潔な文に加工する手法[2]を提案している. 文末表現

に着目しており, 作成したルールに基づいて文を短縮化している. それに対して, 本手法では決算短信を対象としている点と, 深層学習を用いて業績要因文を短縮化している点が異なる.

2. 業績結果の抽出

本手法では業績結果の種類として, 「売上高」, 「営業利益」, 「経常利益」, 「純利益」の4つに限定した. 業績結果について記載している文を抽出する方法として, 文末が「た」の文を対象とした. これは結果について記載している文が「た」で終わることが多いからである. 次に, 業績結果について記載している文で頻出する単語を含んでいる文を対象とした. 頻出する単語として「売上」や「利益」, 「前期」などが存在する. 最後に, 各決算短信において今までの条件で残っている文の中で最初に出現したものを業績結果について記載している文として抽出した. これは決算短信において今回抽出したい全体の業績結果がセグメントごとの業績結果より前に記載されていることを利用した. 抽出した文を利用して, 業績結果の種類と前期からの増減率をペアで抽出し, その中で最も増減率が大きかったものを業績結果とした. 業績結果について記載している文とその文から抽出された業績結果の例をそれぞれ以下に示す.

以上の結果, 当第3四半期連結累計期間における当社グループの売上高は1,571億56百万円(前年同期比14.1%増), 営業利益は28億75百万円(前年同期比57.6%増), 経常利益は29億50百万円(前年同期比50.5%増), 四半期純利益は27億24百万円(前年同期比186.0%増)となりました

純利益 186.0

3. 業績要因の生成

酒井らは決算短信や企業の業績発表記事から業績要因文を抽出する手法を提案している[3][4][5]. これらの手法により決算短信からの業績要因文の抽出を行った. しかし, 抽出した業績要因文は文が長いため, そのままタイトルにおける業績要因として利用することができない. 業績要因文を短くするために, ニューラルネットワークモデルの1つである Seq2Seq(Sequence to Sequence)モデルに注目した. Seq2Seq は RNN(リカレントニューラルネットワーク)を用いたエンコーダーとデコーダーからなるモデルである. シーケンスのペアを学習させることで, 片方のシーケンスを入力するとそれに対応するシーケンスを出力することができる. この性質から翻訳や対話生成, 文章要約で主に利用されている. このモデルを利用することで, 決算短信の業績要因文を入力するとタイトルに含まれているような業績要因が出力されることを試みる. ここで, Seq2Seq モデルの作成は Luong らの手法[6]を参考にして行った.

3.1. 業績要因生成のための学習データの生成

業績要因生成のための学習データとして, 本手法では日経新聞記事集合から抽出される業績発表記事におけるタイトルと, その記事に含まれる業績要因文をペアで抽出し, 学習データとした. 業績発表記事からの業績要因文の抽出方法は酒井らの手法[3]をもとに行った. 以下に抽出したタイトルと業績要因文のペアの例を示す.

<タイトル> 住友不、純利益40%増、4～6月、ビル賃貸事業が好調
<業績要因文> オフィス市況が悪化する中で、主力のビル賃貸事業が好調だった

1つの業績発表記事からは複数の業績要因文が抽出される. そのため, 複数の業績要因文の中からタイトルに採用された業績要因文を推定しペアとする必要がある. この推定処理は, 以降に示す手法にて自動的に行った.

必要な情報はタイトルにおける業績要因と業績要因文なので, それ以外を除去する必要がある. 読点で区切られているタイトルの多くが, 読点で文を分けた時に最後になる節に業績要因について記載していた. このことを利用して, タイトルを読点で分断し, 最後の節のみを抽出した. 以下に例を示す.

<タイトルにおける業績要因> ビル賃貸事業が好調
<業績要因文> オフィス市況が悪化する中で、主力のビル賃貸事業が好調だった

しかし, 作成したデータの中にはタイトルにおける業績要因と業績要因文が別のことについて記載しており, 対応がとれていないペアが存在した. 解決する方法として, タイトルにおける業績要因の最初の単語を重要だと考え, その単語が出現した以降の業績要因文のみを抽出するようにした. 業績要因文がその単語を含んでいない場合は, データから除去した. 以下に例を示す.

<タイトルにおける業績要因> ビル賃貸事業が好調
<業績要因文> ビル賃貸事業が好調だった

なお数字が学習データに含まれていると良いモデルが作成できなかったため, 数字を含んでいないペアのみを学習データとした.

3.2. Seq2Seq による業績要因の生成

作成した学習データを使用して深層学習を行い, 決算短信の業績要因文を入力したところ, ほとんどの入力文に対してエラーが出てしまった. 学習データに含まれていない単語が入力文である決算短信の業績要因文に含まれているのが原因であった. 決算短信にはその企業特有の単語が記載されていることが多く, 業績発表記事にはその単語が含まれていなかった. また, 決算短信は丁寧な言い回し(「しました」など)で記載されていることが多く, 業績発表記事に含まれていない表現が存在した. これらの問題を解決するために, 以下のよう

に学習データを変更した。

- ・文末表現のみを学習データとする。
- ・丁寧な言い回しに変更する。

変更後の学習データの例を以下に示す。

<タイトルにおける業績要因の文末表現> 好調
<業績要因文の文末表現> 好調でした

変更した学習データを用いて、深層学習を行った。決算短信の業績要因文も同様に文末表現のみを抽出して、入力文とした。以下に例を示す。

<入力文>	<出力文>
維持できました	好調
寄与して参りました	寄与
寄与しませんでした	低迷
伸びました	好調
需要不振でした	不振
下回りました	低迷
回復しはじめました	回復

文末表現の生成はできたが、このままでは業績要因にはならない。そのため、決算短信からの業績要因文の抽出の際に利用した企業キーワード[5]に注目した。企業キーワードとは、その企業における重要なキーワードであり、企業ごとの決算短信の集合から抽出される。決算短信の業績要因文における文末表現に係っている単語の中で、企業キーワードのスコアが最も大きいものを抽出した。そして、抽出した単語と、Seq2Seqモデルで生成した文末表現を合成して、決算短信のタイトルに使用する業績要因を生成した。以下に例を示す。

HDD関連部品が好調 半導体デバイス事業が好調 電子デバイス事業が拡大 LEDが寄与 パソコン関連が不振 携帯電話出荷台数が低迷 PC向け製品が上方修正
--

4. タイトルの生成

抽出した業績結果と生成した業績要因を合成してタ

イトルの生成を行う。ここで、業績結果における増減率が増加した場合はポジティブな業績要因、減少している場合はネガティブな業績要因を選択し、増減率と業績要因の極性を一致させる必要がある。そのため、業績要因文の抽出を行った際に獲得した手がかり表現に注目した。手がかり表現には「好調でした」や「低迷しました」など業績がどのような状態だったかを表す表現が多い。「好調でした」ならポジティブ、「低迷しました」ならネガティブなどのように人手で極性を付与した。極性の付与ができないと判断したものは、手がかり表現のリストから除去した。

極性が付与された手がかり表現のリストを用いて業績要因への極性の付与を行う。方法として、業績要因の元となる業績要因文の文末表現を抽出し、手がかり表現のリストの中で一致するものを探す。存在した場合は手がかり表現と同じ極性を付与し、存在しなかった場合はその業績要因をタイトルの候補から除去した。

業績要因への極性付与を行ったが、業績結果の極性と一致する業績要因が複数存在する可能性がある。その場合は業績要因文の抽出を行った際に使用した企業キーワードに注目した。重要な企業キーワードを多く含んでいる業績要因文を使用した方がより良い業績要因を生成できると考えた。候補となっている業績要因の中で、元の業績要因文に含まれている企業キーワードのスコアの合計が最も大きいものをタイトルに使用する業績要因とした。

企業名、業績結果、業績要因を用いて、タイトルの自動生成を行った。生成されたタイトルの例を以下に示す。

- | |
|---|
| <ul style="list-style-type: none">・フューチャー、純利益123.8%増、ITコンサルティング事業が好調・カルビー、売上高21.3%増、シリアル食品が好調・京進、売上高110.0%増、英会話事業が好調・コニカミノルタ、売上高1.4%増、AeroDRが拡大・長谷工コーポレーション、営業利益95.3%増、賃貸マンション運営が好調 |
|---|

5. 評価

自動生成した決算短信のタイトルの評価は、生成された12,422個のタイトルの中からランダムに選択した100個のタイトルを対象にして行った。評価方法は以下の3つの項目を満たしている場合は正解、1つでも満たしていない場合は不正解とした。

- ・業績結果と文末表現の極性が一致しているか。
- ・業績に影響を与えたとされるものが記載されているか。
- ・日本語として自然であるか。

評価の結果、正しいタイトルとして生成できたのが72個、間違っただけのタイトルとして生成されたのが28個であった。よって、正解率は72%であった。不正解となったタイトルの例を以下に示す。

- | |
|---|
| <ul style="list-style-type: none">・昭和真空、売上高 30.1%増、納入が低迷・大倉工業、売上高 0.1%増、販売数量が好調・大同工業、売上高 7.0%増、用チェーンが好調 |
|---|

6. 考察

自動生成されたタイトルについて考察する。生成されたタイトルにおいて不適切なものがいくつかあったが、大きく分けて二つの要因が考えられる。

一つ目は深層学習で生成した文末表現の中で、一部正しく生成できていなかった点が考えられる。学習データの作成を自動的に行っているため、正しくないものを学習してしまっている場合がある。そのため、入力文が正しくないデータに類似していると、正しい出力ができていなかった。作成したデータから人手で正しくないものを除去するなどの改善を行えば、より良い結果が得られると考えられる。

二つ目は文末に係っている単語の中で企業キーワードのスコアが高いものを業績要因に使用した点が考えられる。例えば「販売数量」が業績要因に使用する企業キーワードとして選択されることがあった。本手法の場合、「販売数量」の前に記載されていると考えられる商品名などを選択することができない。企業キーワードのみではなく、企業キーワードに係っている複数の文節を業績

要因に使用するなどの手法を考える必要がある。

最後に評価方法について考察する。今回の評価方法の場合ではタイトルの自然さを基準にして評価したため、生成されたタイトルに含まれる業績要因が最も業績に貢献したか(もしくは不振の主とした要因であったのか等)どうかについては評価できていない。よりタイトルとして厳密な評価をするためには、決算短信の内容を人手にて精査し、生成されたタイトルと一致しているかの確認が必要であるが、この評価のためには多大な労力が必要になるため、今後の課題とする。

7. むすび

本研究では、企業の決算短信から業績結果と業績要因文を抽出し、それらを利用して各決算短信に対してタイトルを生成する手法を提案した。特にタイトルに使用する業績要因については、企業の業績発表記事から学習データを自動的に作成し、Seq2Seqを用いて生成した。評価の結果、正解率は72%であった。

今後の課題として、タイトルの精度を向上させるために、文末表現の生成をする際に使用した学習データの改善をする必要がある。評価方法については、より厳密な評価のために、生成されたタイトルと決算短信の内容が一致しているかを人手にて判定する必要がある。

参考文献

- [1] <https://pr.nikkei.com/qreports-ai/>
- [2] 山本和英, 池田諭史, 大橋一輝: 「新幹線要約」のための文末の整形, 情報処理, Vol. 12, No. 6, pp. 85-111, 2005
- [3] 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
- [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] 酒井浩之, 増山繁: 企業の業績発表記事からの重要業績要因の抽出, 電子情報通信学会論文誌D, Vol. J96-D, No. 11, pp. 2866-2870, 2013
- [6] Minh-Thang Luong, Hieu Pham, Christopher D. Manning: Effective approaches to attention-based neural machine translation, ACL, 2015