

有価証券報告書からの事業セグメントごとの 業績要因文・業績結果文の抽出

Extraction of causal and result information belong to business segments
from securities reports

高野 海斗¹ 酒井 浩之¹ 北島 良三¹

Kaito TAKANO¹, Hiroyuki SAKAI¹, and Ryoza KITAJIMA¹

¹成蹊大学 〒180-8633 東京都武蔵野市吉祥寺北町3-3-1

¹Seikei University 3-3-1 Kichijoji-Kitamachi, Musashino-shi, Tokyo, 180-8633

Abstract: In this research, we propose a method of extracting business segments from securities reports and extracting sentences including causal and result information concerning business performance for each extracted business segments. For example, our method extracts “In the aluminum rolled products business, shipments of high purity foils for aluminum electrolytic capacitors for industrial equipment and automotive use increased and sales increased.” as a sentence including causal information concerning business performance. Moreover, our method estimates that the sentence belongs to business segment “aluminum”. Our method extracts “As a result of this segment sales in this segment were 105,439 million yen, operating profit was 6,697 million yen.” as the result information belong to “aluminum” segment.

1. はじめに

近年、人工知能分野の手法や技術を金融市場における様々な場面に応用することが期待されており、膨大な金融情報を分析して投資判断を支援する技術が注目されている。さらに、最近では証券市場における個人投資家の比重が増大しており、個人投資家に対して投資判断の支援を行う技術の必要性が高まっている[1][2][3]。

投資家が投資活動を行うにあたり、上場企業の業績情報の収集は必要不可欠である。また、業績情報の中でも特に業績要因が投資判断において重要である。なぜなら、業績回復の要因が、その企業の主力事業が好調であることであったならば株価への影響は大きい、株式売却益の計上などの特別利益の計上が要因であるならば株価への影響は軽微であるからである。

これまで、決算短信から業績要因の抽出は行われてきた[4][5]。しかし、決算短信は、上場企業が決算発表時に提出する決算速報であるため、タイムリーな情報が得られる反面、企業によってフォーマットが異なっていることが問題となる。特に、事業セグメントごとの業績情報は、企業ごとに表や文書のフォーマットが異なるため、事業セグメント名や事業セグメントごとの業績情報の抽出は困難である。

そこで本研究では有価証券報告書から事業セグメントごとの業績要因と業績結果の抽出を行う。有価証券報告書は、金融商品取引法で規定されており、審査を通す必要があることから、公表までに時間がかかる反面、決算短信と比較するとフォーマットが定まっている。また有価証券報告書のデータは、金融庁が運営している Web ページである EDINET[8]上に開示されているため、収集が容易である。

本研究では、事業セグメントごとの業績要因文を抽出し、さらに、業績結果と組み合わせることで、抽出された業績要因文がその企業の中でどの程度の重要性をもつかを判断するのに役立つと考える。例えば、株式会社 SUBARU では事業セグメントが「自動車」と「航空宇宙」の2つある。それぞれのセグメントごとの2016年3月31日から2017年4月1日までの利益は、図1.1からわかりように、「自動車セグメント」で397,657百万円、「航空宇宙セグメント」で9,102百万円となっており、大きな差がある。そのため、SUBARUにとって「自動車セグメント」の業績要因は「航空宇宙セグメント」の業績要因よりも重要であると考えられる。

当初は、事業セグメントごとの業績要因文と業績結果文が記載されている箇所（文の集合）を推定してから、業績要因文であるかどうかを分類しようと試みたが、企業や年度によってフォーマットが異な

るため、適合率・再現率に限界があった。そこで、本手法では、先に業績要因文であるかどうかの判定を行い、そこに事業セグメントを付与する。

関連研究として、村野らは企業の決算短信から自動抽出した業績要因文に対して、事業セグメントをk近傍法によって分類している[6]。しかし、村野らの手法を行うためには、人手で事業セグメント名を企業Webページや、決算短信から収集する必要がある。証券市場の上場企業数は東京証券取引所の上場企業だけでも2018年10月1日現在、3,635社も存在するため、人手で付与するには多大な労力が必要となる。それに対し、本研究で提案する手法は、ある程度フォーマットが定まっている有価証券報告書を用いることで、事業セグメント名を自動的に抽出し、人手による作業を必要としないものとなる。

	売上高				セグメント利益			
	2016年3月期	2017年3月期	増減	増減率	2016年3月期	2017年3月期	増減	増減率
自動車	3,039,424	3,151,961	112,537	3.7	543,609	397,657	△145,952	△26.8
航空宇宙	152,786	138,759	△14,027	△9.2	18,201	9,102	△9,099	△50.0
その他	40,048	35,272	△4,776	△11.9	2,998	3,512	514	17.1
調整額	-	-	-	-	781	539	△242	△31.0
合計	3,232,258	3,325,992	93,734	2.9	565,589	410,810	△154,779	△27.4

(注) 1. 売上高は、外部顧客への売上高であります。
2. セグメント利益の調整額は、セグメント間取引消去であります。

図 1.1. 株式会社 SUBARU H29 セグメント状況

2. 提案手法

本研究では、有価証券報告書から企業ごとの事業セグメント名とその事業セグメントに関する業績要因文・業績結果文を抽出する手法を提案する。以下に本手法の概要を示す。

- Step 1: 有価証券報告書の 1.5 節「従業員の状況」から、事業セグメント名の候補を抽出する。
- Step 2: 有価証券報告書の 2 章「事業の状況」から、文をすべて抽出し、業績要因文、業績結果文、それ以外の文に分類する。
- Step 3: Step 2 で抽出した業績要因文に対して、事業セグメントを付与し、その業績要因文に対応する業績結果文があれば、その業績結果文にも事業セグメントを付与する。
- Step 4: Step 3 で事業セグメントが付与できなかった業績要因文に対して、別のアプローチを用いて、事業セグメントを付与する。

2.1. 事業セグメント名候補の抽出

有価証券報告書の 1.5 節「従業員の状況」に事業セグメントごとの従業員数が記載されているため、この部分から事業セグメント名と思われる文字列を候補として抽出する。具体的には表中の「セグメント

の名称」と「合計」の間にある文字列を事業セグメント名の候補とする。

図 2.1 は昭和電工の有価証券報告書からの事業セグメント名候補抽出の例である。昭和電工は「石油化学」、「化学品」、「エレクトロニクス」などが事業セグメント名であり、それらが候補として抽出される。

(2) 親出会社の状況			
平成29年12月31日現在			
従業員数(名)	平均年齢(歳)	平均勤続年数(年)	平均年間給与(円)
3,616 (821)	40.0	16.6	7,185,227

セグメントの名称	従業員数(名)
石油化学	406 (71)
化学品	1,234 (172)
エレクトロニクス	321 (127)
無属	251 (74)
アルミニウム	623 (271)
その他	781 (106)
合計	3,616 (821)

(注) 1 従業員数は就業人員であり、社外への出向者を除き、社外から受け入れた出向者を含む。また、執行役員及びコーポレートフェローを含まない。
2 臨時雇用者数(契約社員、嘱託社員を含む。)は、当事業年度の平均人員を() 外数で記載している。
3 平均年間給与は、賞与及び基準外賃金を含んでいる。
4 全社共通研究に係る従業員については、「その他」に含めて表示している。

(3) 労働組合の状況
当社は、昭和電工会ニオンがあり、本部を東京都港区に置いている。(平成29年12月31日現在組合員数2,832名。出向中の組合員を含まない。)
会社と昭和電工会ニオンとの間には、相対的平和義務を伴う労働協約が締結されており、友好的な関係を維持している。
連結子会社における労働組合の状況について特に記載すべき事項はない。

- 9 -

図 2.1. 事業セグメント名候補抽出の例

2.2. 業績要因文、業績結果文の抽出

有価証券報告書の 2 章「事業の状況」から文の集合を抽出する。次に、抽出した文集合を一文ずつ分類し、業績要因文と業績結果文を抽出する。分類は文が業績要因文であるかどうかの 2 値分類と、業績結果文であるかどうかの 2 値分類を異なる手法を用いて行う。

①業績要因文であるかどうかの分類

業績要因文の分類は深層学習を用いて行った。深層学習のための学習データは酒井らの手法[7]により自動的に生成した。酒井らの手法の概要と、業績要因文の判定は以下のとおりである。

- Step 1: 酒井らの手法[5]を用いて、決算短信から業績要因文、手がかり表現、後述の企業キーワードを抽出する。
- Step 2: Step 1 で抽出された手がかり表現の“拡張手がかり表現”を獲得する。
- Step 3: Step 1 で抽出された業績要因文に対して、企業キーワードを用いてスコアを付与する。
- Step 4: 拡張手がかり表現を含み、かつ、スコアが高い業績要因文を正例とし、手がかり表現、企業

キーワードをともに含まない文を負例として学習データを自動生成する。

Step 5: 自動生成された学習データを使用し、深層学習にて有価証券報告書から抽出した事業セグメントに対応する文集合に対して、1文ごとに業績要因文を判定する

ここで、企業キーワードとはその企業にとって重要なキーワードのことを示しており、以下の式を用いて、名詞に重み付けを行い、重みが大きい名詞を対象とする。

$$W(n_i, S(t)) = Tf(n_i, S(t)) \times H(n_i, S(t)) \times idf(n_i)$$

$S(t)$: 企業 t の決算短信の集合。

$Tf(n_i, S(t))$: $S(t)$ において名詞 n_i が出現する頻度。

$H(n_i, S(t))$: $S(t)$ の要素である決算短信に名詞 n_i が出現する確率に基づくエントロピー。

深層学習のモデルは多層パーセプトロンを採用し、隠れ層は、順にノード数 3,000 が 3 層、ノード数 1,000 が 3 層、ノード数 500 が 3 層、ノード数 100 が 3 層の計 12 層とした。出力層は 1 要素である。また、epoch 数は 20、活性化関数として ReLU を使用した。

さらに、深層学習によって業績要因文と判定された文集合から、上述の企業キーワード抽出のための名詞の重み付けの式を用いて企業キーワードを新たに抽出し、業績要因文は新たに抽出した企業キーワードを含むものに限定した。

②業績結果文であるかどうかの分類

ルールベースで「億円」「万円」「千円」などを含む文とし分類を行った。ただし、このルールだけでは業績結果文ではない文も多く業績結果文として誤分類してしまうため、2.3 節で述べる事業セグメントの付与において、業績要因文との関係に基づき業績結果文を確定させる。

2.3. 事業セグメントの付与

2.2 節で抽出した業績要因文と業績結果文に対して、2.1 節で獲得した事業セグメント名の候補を用いて、事業セグメントの付与を行う。付与の手法は以下のとおりである。

① 文の位置による付与

有価証券報告書の 2 章「事業の状況」を上から 1 文ごとに取得する。「。」を含まない文であれば、事業セグメント名の候補を含んでいるかどうか確認する。事業セグメント名候補を含んでいれば、そこか

ら 5 行以内に出てきた文は、その事業セグメントの内容であると判定する。5 行以内に業績要因文や業績結果文と判定された文が取得された場合、そこから新たに 5 行以内に出てきた文は、その事業セグメントの内容であると判定する。

具体的には、表 2.2 の例だと、26 行目の行が「。」を含まない文であり、「コンサルティング」が事業セグメント名候補として抽出できているため、以下 5 行の 31 行目までは「コンサルティング」事業の文であると判定する。31 行目の文は業績結果文であるため、さらに以下 5 行の 36 行目までは「コンサルティング」事業の文であると判定される。しかし、32 行目の行が「。」を含まない文であり、「金融 IT ソリューション」が事業セグメント名候補として抽出できているため、32 行目から以下 5 行の 37 行目までは、「金融 IT ソリューション」事業の文であると判定が上書きされる。35 行目の文は業績結果文であるため、さらに以下 5 行の 40 行目までは「金融 IT ソリューション」事業の文であると判定される。

手法①での事業セグメントの付与は、適合率が高いものの、フォーマットの関係などで事業セグメントが付与できていない業績要因文が多く存在する。そこで、①で付与できなかった業績要因文に対して以下のアプローチで事業セグメントの付与を行った。

② ルールベースによる付与

まず、第一段階として、xbrl ファイルからテキストを抽出する際に、事業セグメント名と事業セグメントに対応する文が一つになってしまい、事業セグメントの付与ができない原因に対応した。

例えば、「アルミニウム」、「当セグメントでは、アルミ電解コンデンサ…」のように行が分かれてテキストを抽出することを想定しているが、この抽出が「アルミニウム当セグメントでは、アルミ電解コンデンサ…」のように 1 行になって抽出してしまうことが起こっている。そこで、行の先頭、1 文字削除した箇所、2 文字削除した箇所、3 文字削除した箇所に対して、事業セグメント名の候補があるかどうかでセグメントの付与を行った。1 文字～3 文字削除した箇所に関しては、「①アルミニウム当セグメント…」や「(1)アルミニウム当セグメント…」に対応するためである。これにより、新たに 20621 の業績要因文に事業セグメントを付与することができた。

表 2.2. 野村総合研究所の業績要因文・業績結果文へのセグメント付与例

行数	業績要因	業績結果
26		(コンサルティング)
27		当セグメントは、政策提言や戦略コンサルティング、業務改革をサポートする業務コンサルティング、ITマネジメント全般にわたるシステムコンサルティングを提供しています。
28		顧客の経営環境やIT部門の環境が変化中、経営・ITの両面でコンサルティングの需要が高まっています。
29	○	当社グループは、顧客のビジネス全般を支援する変革パートナーとなる体制を整えていくとともに、海外も含めた顧客基盤の拡大に努めました。
30	○	当年度は、企業収益の改善を受け、顧客業務の実行を支援する業務コンサルティングなどが増加したことに加え、グローバル関連ではASGGroupLimitedがシステムコンサルティングの増加に寄与しました。
31	○	この結果、売上高31,161百万円(前年度比8.1%増)、営業利益5,853百万円(同6.7%増)となりました。
32		(金融ITソリューション)
33		当セグメントは、主に証券業や保険業、銀行業等の金融業顧客向けに、システムコンサルティング、システム開発及び運用サービスの提供、共同利用型システム等のITソリューションの提供を行っています。
34		事業領域の拡大に向け、業界標準ビジネスプラットフォームの生産革新を進めるとともに、ITと金融を融合したFinTech(フィンテック)等を活用した新事業の開発に取り組んでいます。
35	○	既存事業の拡大に向けた取組みとして、業界標準ビジネスプラットフォームについては、制度改正への着実な対応を進めるとともに、顧客業務の高度化や効率化の支援を進めました。
36		リテール証券のバックオフィス業務をサポートする共同利用型システム「STAR-IV」については、災害時にシステム障害復旧を行う機能を追加したサービスを提供しています。
37		また、平成30年に実施予定の国債の決済期間短縮化や、証券保管振替機構の次期システムへの移行について、当社の共同利用型システムの対応を進めています。
38	○	また、資産運用領域の事業拡大を目的に、米国のCutterAssociates,LLCを子会社としました。
39	○	○ 当年度の売上高は、証券業を中心にシステムコンサルティングが増加しましたが、前年度に大型の製品販売があった証券業向け開発・製品販売や、保険業向け開発・製品販売が減少し、248,188百万円(前年度比2.2%減)となりました。
40	○	○ 子会社の(株)だいこう証券ビジネスにおいて業務体制見直しに向けた事業構造改善費用を計上したこともあり、営業利益は26,461百万円(同9.3%減)となりました。

③最近傍法による分類

ここまでに事業セグメントが付与できた業績要因文を用いて、まだ事業セグメントが付与できていない業績要因文とのコサイン類似度を計算し、コサイン類似度が一番高い業績要因文と同様の事業セグメントを付与した。コサイン類似度の計算は、業績要因文から新たに獲得した企業キーワードを要素、企業キーワードの重みを要素値とした単語ベクトルで計算した。

これにより、新たに 42375 の業績要因文に事業セグメントを付与することができた。

3. 実装

本手法を実装し、キーワードによって業績要因文を検索できる有価証券報告書検索システムを作成した¹。使用するデータは EDINET から取得した有価証券報告書の xbrl ファイル (21,324 個) を用いた。

図 3.1 に「燃料電池」で検索した場合の検索結果を示す。アルミニウムを含む業績要因文と企業名、その業績要因文に付与された事業セグメント名、さらに、その事業セグメントの業績文が表示される。



図 3.1 「燃料電池」での検索結果

4. 評価

業績結果文であるかどうかの分類精度の評価は、事業セグメントを 5 つ以上持つ企業をランダムに 10 社、選択し、選択した 10 社の有価証券報告書から抽出された業績要因文 83 文を人手にて正解を判定して精度を計算した。また、事業セグメント付与の精

¹ <http://hawk.ci.seikei.ac.jp/u-cees/>

度は、業績要因文の判定が正しかったもののうち、正しい事業セグメントに分類されているかどうかで評価を行った。その結果を表 4.1 に示す。

表 4.1. 評価結果

企業名	業績要因文の分類精度	事業セグメント付与の精度1	事業セグメント付与の精度2
太洋工業株式会社	1.00	1.00	0.56
株式会社ウィルグループ	0.86	1.00	1.00
東京都競馬株式会社	0.89	1.00	0.88
日本プロセス株式会社	0.88	1.00	0.86
株式会社ハウズドゥ	0.88	1.00	0.71
ユアサ商事株式会社	0.91	1.00	0.70
パーソルホールディングス株式会社	1.00	1.00	0.80
東洋インキSCホールディングス株式会社	1.00	1.00	1.00
ナカバヤシ株式会社	1.00	1.00	1.00
九州旅客鉄道株式会社	0.70	1.00	0.71
平均	0.91	1.00	0.82

事業セグメント付与の精度 1 と 2 の違いは、「鏡面研磨機事業の販売は増加したものの、電子基板事業におけるディスプレイメーカー向けの受注の減少、並びに商社事業及び基板検査機事業の販売が減少したことから、売上高が減少いたしました。」のような企業全体の業績要因文に対して、文に含まれているどこかの事業セグメントが付与されているものを、分類が合っているとしたものが精度 1 であり、分類が間違っているとしたものが精度 2 である。

5. 考察と今後の方針

最近傍法を用いることで、事業セグメントの付与ができた文がある反面、企業全体のことについて記載されている部分に事業セグメントが付与されてしまった。今後は企業全体のことについて記載されているかどうかを推定する必要もある。

業績要因文の分類精度は高い結果となった。誤った業績要因文については、「合意しました」などの文が含まれていた。再現率はまだ測定していないため、今後正解データを作成し測定していく。

事業セグメントごとの業績要因文と業績結果文は、有価証券報告書にまとまって出現するため、今後は事業セグメントごとの記述が記載されている部分を推定していくことで、フォーマットが多少異なっている企業に対しても、抽出ができるようにしていきたい。また、本手法で取得できた事業セグメント名を用いることで、決算短信で抽出した業績要因文に対しても、事業セグメント名を付与できる可能性が高い。

今回自動で取得したデータは、業績要因文と業績結果文であるが、業績要因文の要因には、天候、原材料の価値変化、社会システムの変更、商品需要の

変化など様々な要因があるため、それらを分析することにも価値があると考えられる。また、業績結果文は売上や営業利益等を数値データに変換し、時系列での解析や、全体の売上や営業利益との割合で、その業績要因の重要度を定量的に推定することも可能である。今後は取得したデータの解析も行い、その結果をアウトプットできるようにする予定である。また、有価証券報告書には業績以外にも「事業の内容」、「企業の対処すべき課題」、「事業等のリスク」、「研究開発活動」など企業に関する重要な情報が多く記載されているため、これらを自動で抽出することで、様々な解析を行っていく予定である。

6. むすび

本研究では、これまで決算短信からは抽出することができなかった事業セグメントごとの業績要因文と業績結果文の抽出を有価証券報告書から行った。その結果、業績要因文の分類精度は 0.91 であり、事業セグメントの付与も良好であった。

参考文献

- [1] 和泉 潔, 後藤 卓, 松井藤五郎, “経済テキスト情報を用いた長期的な市場動向推定”, 情処学論, vol.52, no.12, pp.3309-3315, (2011)
- [2] 蔵本貴久, 和泉 潔, 吉村 忍, 石田智也, 中嶋啓浩, 松井 藤五郎, 吉田 稔, 中川裕志, “新聞記事のテキストマイニングによる長期市場動向の分析”, 人工知能学会論文誌, vol.28, no.3, pp.291-296, (2013)
- [3] 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”, 信学論 (D), vol.J96-D, no.11, pp.2866- 2870, (2013)
- [4] 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, vol.J98-D, no.5, pp.811-822, (2015)
- [5] 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)
- [6] 村野壮人, 酒井浩之, 坂地泰紀, 江口潤一, “決算短信から抽出した業績要因文の事業セグメントに基づく分類と業績文の抽出企業の決算短信 PDF からの業績要因の抽出”, 第 19 回金融情報学研究会, pp.59-64, (2017)
- [7] 酒井浩之, 松下和暉, “決算短信からの業績要因文の抽出”, 第 11 回テキストアナリティクス・シンポジウム, pp.87-91, (2017)
- [8] EDINET, <http://disclosure.edinet-fsa.go.jp>