

深層学習と拡張手がかり表現による業績要因文への極性付与 Assigning Polarity to Causal Information by Deep Learning and Extended Clue Expressions

酒井浩之^{1*} 坂地泰紀¹ 山内浩嗣^{2†} 町田亮介^{2†} 阿部一也^{2†}
Hiroyuki Sakai¹ Hiroki Sakaji¹ Hiroaki Yamauchi² Ryosuke Machida² Kazuya Abe²

¹ 成蹊大学 理工学部 情報科学科

¹ Department of Computer and Information Science, Faculty of Science and Technology,
Seikei University

² 三菱UFJトラスト投資工学研究所

² Mitsubishi UFJ Trust Investment Technology Institute

Abstract: In this paper, we propose a method of assigning polarity to causal information extracted from summary of financial statements of companies. Our method assigns polarity (positive or negative) to causal information in accordance with business performance, e.g. “Orders of semiconductor manufacturing equipments were strong”. First, we assign polarity to extended clue expressions to be used to extract causal information. Using them, our method automatically generates training data and assigns polarity to causal information by deep learning. We evaluated our method and confirmed that it attained 86.7% precision and 95.4% recall of assigning polarity positive, and 90.0% precision and 73.9% recall of assigning polarity negative, respectively.

1 はじめに

近年、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おこなう技術の必要性が高まっている。そのため、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、例えば、膨大な金融情報を分析して投資判断の支援を行う技術が注目されている。その一例として、日本銀行が毎月発行している「金融経済月報」や経済新聞記事をテキストマイニングの技術を用いて解析し、経済市場を分析する研究などが盛んに行われている [1]。

投資家にとって、企業の業績に関する情報を収集することは重要であるが、実際の業績に関する情報だけでなく、その業績要因も重要である [4][5]。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。しかしながら、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 2017 年 2 月 17 日現在、3,537 社と多いうえに¹、近年では年に 4 回、決算発表がある。さらに、大幅な業績の修正を行う場合にも業績修正発表を行う必要があるため、人手によって多くの企業の業績要因を取得するには多大な労力を要する。そのために酒井らは、企業が業績発表を行った直後に企業の Web サイト等で公開される決算短信 PDF に着目し、その中から業績要因を含む文（例えば「半導体製造装置の受注が好調でした。」）を抽出する手法を提案した [6]。

決算短信より抽出された業績要因を含む文は、例えば、証券アナリストへの支援材料として利用できる。（以降、業績要因を含む文を「業績要因文」と定義する。）業績要因文の応用例として、抽出された業績要因文を検索対象とし、例えば「太陽電池」で検索することで、「太陽電池」を含む業績要因文（例：「太陽電池は、国内の住宅用及びメガソーラーなど産業用の販売が伸長したことから、売上高は前年同期比 85.8% 増の 2,768 億円となりました。」）を検索することができ、そのような業績要因文が頻繁に出現する企業を検索することができる。本例の「太陽電池」では、「東洋炭素」、「フジプレアム」、「シャープ」、「カネカ」といった企業が検索される。しかし、より有効な情報として利用するためには、抽出した業績要因文に対して

*連絡先：成蹊大学 理工学部 情報科学科
〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1
E-mail: h-sakai@st.seikei.ac.jp

†本稿の内容は筆者が所属する組織を代表するものではなく、すべて個人的な見解である。また、当然のことながら、本稿における誤りは全て筆者の責に帰するものである。

¹<http://www.jpx.co.jp/listing/stocks/co/index.html>

業績に関する極性（「ポジティブ」、「ネガティブ」）を付与する必要がある。例えば、業績要因文「半導体製造装置の受注が好調でした。」に対しては「ポジティブ」、
「世界的な太陽電池市況の低迷により太陽電池製造装置の販売が減少しました。」に対しては「ネガティブ」の極性を付与する。業績要因文に対して極性を付与することで、業績要因を使用した景気動向予測、および、業績要因に基づいて株取り引きを行うコンピュータトレーディングにも応用することが期待できる。例えば、「太陽電池」で検索された業績要因文に極性が付与されていることで、太陽電池の市況を業績要因文に付与された極性の数で得ることができ、また、その決算短信PDFの発表日と組み合わせることで、太陽電池の市況の時系列推移を得ることができる。そのため、本稿では、決算短信PDFより抽出された業績要因文に対して極性（「ポジティブ」、「ネガティブ」）を自動的に付与する手法を提案する。具体的には、業績要因文を抽出するための手がかりとなる表現（以降、手がかり表現と定義）にいくつかの文節を追加することで、“拡張手がかり表現”と定義する文節列（例えば「受注が好調でした」）を獲得する。そして、“拡張手がかり表現”を使用することで極性付与のための学習データを自動的に生成し、生成された学習データを使用して深層学習を行い、業績要因文に極性を付与する。

2 決算短信から抽出した業績要因文への極性付与

まずは抽出した手がかり表現²に極性（ポジティブ、ネガティブ）を手で付与し、その手がかり表現の極性を使用して業績要因文への極性付与を行うことを試みる。しかしながら、例えば「推移しました」のような手がかり表現には極性を付与できない。なぜなら「堅調に推移しました」はポジティブとなるが、「厳しい状況で推移しました」はネガティブとなるからである。このように、手がかり表現に係る文節列（上記の例では「堅調に」「厳しい状況で」）によって極性が変わる手がかり表現が存在する。また、「増加」「増え」「減少」を含む手がかり表現は、その手がかり表現に係っている文節列によって極性が反転する場合がある。例えば、手がかり表現の「増加しました」はポジティブの極性を付与するが、「費用が」や「コストが」が係っている場合はネガティブの極性を付与する必要がある。逆に、手がかり表現の「減少しました」はネガティブの極性を付与するが、「費用が」や「コストが」が係っている場合はポジティブの極性を付与する必要がある。このように、極性を付与できない手がかり表現を含む業績

²172,951 個の決算短信 PDF から酒井らの手法 [6] により 163 種類の手がかり表現を抽出した。

表 1: 極性付与された手がかり表現の例

手がかり表現	極性
堅調でした	positive
伸び悩んだ	negative
寄与し	positive
低調でした	negative
低迷し	negative
改善し	positive

要因文には極性を付与できず、さらに、極性を付与した手がかり表現のみでは業績要因文の極性を決定することができない場合がある。そのため、あらかじめ極性を付与した業績要因文を大量に作成し、それを学習データとした深層学習により、業績要因文に極性を付与する。しかしながら、あらかじめ極性を付与した業績要因文を手で大量に用意することは、多大な労力を要する。そこで本研究では、業績要因文への極性を付与するための学習データを自動的に生成する。

まず、抽出した手がかり表現のなかから、極性（ポジティブ、ネガティブ）を付与できる手がかり表現に対して手で極性を付与する。その結果、163 種類の手がかり表現のうち、108 種類の手がかり表現に極性を付与した。表 1 に、手で極性付与した手がかり表現の一部を示す。しかし、極性付与した手がかり表現を含む業績要因文を学習データとして全て使うことは適切ではない。なぜなら、業績要因文の抽出精度が 83% 程度であり、誤った業績要因文も多く含まれているからである。そこで、極性が付与できる手がかり表現にいくつかの文節を追加することで、“拡張手がかり表現”と定義する文節列（例えば「受注が好調でした」）を獲得する。これにより、例えば手がかり表現「好調でした」が「受注が好調でした」や「極めて好調でした」のような、より精度の高い手がかり表現へ拡張される。そして、例えば拡張手がかり表現「極めて好調でした」を含む業績要因文をポジティブ、拡張手がかり表現「極度の不振で」を含む業績要因文をネガティブとすることで学習データを自動的に生成する。

以上の説明をふまえ、本手法の概要を以下に示す。

Step 1: 決算短信 PDF から抽出した業績要因文から、極性を付与できる手がかり表現の“拡張手がかり表現”を獲得する

Step 2: ポジティブの拡張手がかり表現を含む業績要因文をポジティブ、ネガティブの拡張手がかり表現を含む業績要因文をネガティブとして、学習データを自動生成する。

Step 3: 拡張手がかり表現を含まない業績要因文に対して、Step 2 で生成された学習データを使用し、

深層学習にて極性を付与する。

2.1 拡張手がかり表現の獲得

極性を付与した手がかり表現に文節列を追加して、拡張手がかり表現を獲得する手法について述べる。具体的には、手がかり表現 c に係る文節列 p に対して以下の式 1 でスコアを求め、このスコアがある閾値を上回る文節列を抽出する。

$$\begin{aligned} \text{Score}(p, c) &= -f(p, c)\sqrt{fp(p)}\log_2 P(p, c)H(p) \\ P(p, c) &= \frac{f(p, c)}{N(c)} \end{aligned} \quad (2)$$

ただし、決算短信 PDF から取得した業績要因文の集合において、

$P(p, c)$: 手がかり表現 c から取得される文節列 p の出現確率

$f(p, c)$: 手がかり表現 c から取得される文節列 p の取得回数

$N(c)$: 手がかり表現 c から取得される文節列の総数

$fp(p)$: 文節列 p に含まれる文節の数,

$H(p)$: 文節列 p がある企業の業績要因文に出現する確率に基づくエントロピー (後述)。

例えば、決算短信 A に「収益性が大幅に改善しました」という業績要因文が存在していたとすれば、手がかり表現「改善し」に係る文節列として「大幅に」、「収益性が大幅に」という 2 つの文節列を取得する。また、決算短信 B に「業績が大幅に改善しました」という文が存在していたとすれば、この文から「大幅に」、「業績が大幅に」という 2 つの文節列を取得する。そして、決算短信 A と決算短信 B からは「大幅に」が 2 回、「収益性が大幅に」、「業績が大幅に」をそれぞれ 1 回、取得したことになる。そのため、例えば「大幅に」の $f(p, c)$ の値は 2 であり、 c から取得する文節列の総数 $N(c)$ は 4 であるため、 $P(p, c)$ の値は $2/4$ となる。

$H(p)$ は文節列 p がある企業の業績要因文に出現する確率に基づくエントロピーであり、 $H(p)$ が高い文節列は多くの企業の業績要因文に出現している文節列であることが分かる。また、1 つの企業にのみ多く出現する文節列の $H(p)$ は 0 になるため、そのような文節列を除去できる。 $H(p)$ は以下の式 3 で求める。

$$H(p) = - \sum_{s \in S(p)} P(p, s) \log_2 P(p, s) \quad (3)$$

ここで、 $S(p)$ は文節列 p を含む業績要因文をもつ企業の集合、 $P(p, s)$ は文節列 p が企業 s の業績要因文に出現する確率を表す。

手がかり表現 c における $\text{Score}(p, c)$ の平均を求め、平均値より大きいスコアが付与された文節列 p を手がかり表現 c に追加し、それを拡張手がかり表現として獲得する。表 2 に 172,950 の決算短信 PDF から抽出された業績要因文を対象に、手がかり表現「改善し」から取得された文節列の中で $\text{Score}(p, c)$ が高い文節列をいくつか示す。表 2 に示すように、手がかり表現に追

表 2: 「改善し」から取得された文節列

文節列	$\text{Score}(p, c)$	拡張手がかり表現
大幅に	18,543	大幅に改善し
大きく	10,157	大きく改善し
利益率が	5,438	利益率が改善し
収益が	5,278	収益が改善し
採算が	3,066	採算が改善し

加する文節列は 1 文節であることが多いが、拡張手がかり表現が「企業の景況感が改善する」「当初の計画を上回って」「売上高の増加に貢献した」のように、複数の文節で構成される文節列が追加される場合もある。

2.2 拡張手がかり表現の極性

拡張手がかり表現の極性は、拡張前の手がかり表現の極性に準じる。例えば拡張手がかり表現「大幅に改善し」の場合、拡張前の手がかり表現「改善し」に付与された極性がポジティブであるため、ポジティブとなる。ただし、「増加」、「増え」、「減少」を含む手がかり表現は、拡張手がかり表現によっては極性が反転する場合がある。例えば手がかり表現「増加する」はポジティブの極性が付与されており、その拡張手がかり表現「売上高が増加する」はポジティブであるが、「人件費が増加する」はネガティブの極性を付与する必要がある。そのため、「増加」、「増え」、「減少」を含む手がかり表現に以下に示す語を含む文節列が追加されている場合、手がかり表現に付与されている極性の逆の極性を、拡張手がかり表現に付与する。

費, 支出, 借入金, 債権, 未払金, 債務, 負債, 損失, コスト, 貸出金

極性を付与した 108 種類の手がかり表現から、4,318 個のポジティブの拡張手がかり表現、1,898 個のネガティブの拡張手がかり表現を獲得した。

2.3 学習データの自動生成

拡張手がかり表現とその極性を使用して、学習データを自動生成する。ここで、業績要因文には例として以下に示すように複数の単文で構成される複文が含まれているとする。

粉ミルクが震災直後のまとめ買いの反動などによる需要減少により前年同期を下回りましたが、流動食は生産・供給能力回復によりほぼ前年同期並みとなりました

このような業績要因文の場合、接続助詞「が」で分割し、「粉ミルクが震災直後のまとめ買いの反動などによる需要減少により前年同期を下回りました」と「流動食は生産・供給能力回復によりほぼ前年同期並みとなりました」のそれぞれに対して極性を付与する必要があります。また、学習データもそれに準拠して生成する必要があります。そのため、業績要因文の文中に接続助詞「が」、および、接続助詞「ものの」が出現したら、そこで業績要因文を区切る。以降、接続助詞によって区切られた文も含めて業績要因文とする。

業績要因文に拡張手がかり表現が含まれていれば、その拡張手がかり表現の極性を業績要因文に付与して学習データとする。以下に自動生成された学習データの一部を示す。+1が正例（ポジティブ）、-1が負例（ネガティブ）であることを示す。また、下線は含まれていた拡張手がかり表現を示す。

+1 血液凝固阻止剤などの販売量が大幅に増加しました
+1 イーファイナンス事業の売上高の増加も大きく貢献しました
-1 健康機能性飲料が前年同期の売上げを大きく下回ったこと等により減収となりました

後述の評価では、340社を無作為に選択し、その業績要因文から47,953文の学習データを自動生成した。

3 深層学習による業績要因文への極性付与

3.1 素性選択

自動生成された学習データに基づき、深層学習により業績要因文へ極性を付与する。まず、入力層の要素となる語（素性）を選択する。具体的には、自動生成された学習データにおいて正例の業績要因文に含まれる内容語（名詞、動詞、形容詞）に対して、以下の式4にて重みを計算する。

$$W_p(t, S_p) = TF(t, S_p)H(t, S_p) \quad (4)$$

ただし、

S_p : 学習データにおいて正例に属する業績要因文の集合

$TF(t, S_p)$: 文集合 S_p において、語 t が出現する頻度

$H(t, S_p)$: 文集合 S_p における各業績要因文に含まれる語 t の出現確率に基づくエントロピー

$H(t, S_p)$ が高い語ほど、正例の文集合に均一に分布している語であることが分かる。 $H(t, S_p)$ は次の式5で求める。

$$H(t, S_p) = - \sum_{s \in S_p} P(t, s) \log_2 P(t, s) \quad (5)$$

$$P(t, s) = \frac{tf(t, s)}{\sum_{s \in S_p} tf(t, s)} \quad (6)$$

ここで、 $P(t, s)$ は業績要因文 s における語 t の出現確率を表し、 $tf(t, s)$ は文 s において語 t が出現する頻度を表す。次に、負例の業績要因文に含まれる内容語（名詞、動詞、形容詞）に対しても、同様に重みを計算する。

$$W_n(t, S_n) = TF(t, S_n)H(t, S_n) \quad (7)$$

ただし、 S_n は学習データにおいて負例に属する業績要因文の集合である。

ここで、ある語 t の正例における重み $W_p(t, S_p)$ が負例における重み $W_n(t, S_n)$ の2倍より大きければ、その語 t を素性として選択する。もしくは、語 t の負例における重み $W_n(t, S_n)$ が正例における重み $W_p(t, S_p)$ の2倍より大きければ、その語 t を素性として選択する。すなわち、以下の条件のどちらかが成り立つ語 t を素性として選択する。

$$W_p(t, S_p) > 2W_n(t, S_n) \quad (8)$$

$$W_n(t, S_n) > 2W_p(t, S_p) \quad (9)$$

上記の条件を課すことで、正例、負例における特徴的な語のみを素性として選択し、正例、負例、ともによく出現するような一般的な語を素性から除去する。以下に選択された素性の一部を例示する。

増加, 減少, 好調, 推移, 拡大, 利益, 影響, 大幅, 増収, 伸長, 順調, 減収

後述の評価における47,953文の学習データでは、5,920語が素性として選択された。

3.2 モデル

入力は、47,953文の学習データから抽出された5,920語を要素、語 t における $\log(W_p(t, S_p))$ 、もしくは、 $\log(W_n(t, S_n))$ の大きいほうを要素値としたベクトルとする。モデルの入力層のノード数を入力ベクトルの次元数と同じ5,920とし、隠れ層は、ノード数1,000が3層、ノード数500が3層、ノード数200が3層、ノード数100が3層の計12層とする。出力層は1要素である。

表 3: 業績要因文への極性付与の評価結果 (本手法)

	positive	negative
精度 (%)	86.7 (210/242)	90.0 (91/101)
再現率 (%)	95.4 (210/220)	73.9 (91/123)

4 評価

本手法の評価を行うため、本手法を実装した。実装にあたり、形態素解析器として MeCab³、係り受け解析器として CaboCha[3] を使用した。ここで、340 社を無作為に選択し、その業績要因文から 47,953 文の学習データを自動生成した。テストデータは、学習データには含まれない (すなわち拡張手がかり表現を含まない) 業績要因文から無作為に 343 文の業績要因文を抽出して作成した。テストデータにおいて本手法によりポジティブの極性が付与された業績要因文を以下にいくつか示す。

- ・豪州の鉄鉱石事業が堅調に推移し、トレード関連では原油・石油製品取引及び鉄鋼製品取引も堅調に推移しました
- ・検索連動型広告においては、スマートフォン経由の売上が大きく伸び、売上が前年同四半期比で増加しました

また、テストデータにおいて本手法によりネガティブの極性が付与された業績要因文を以下にいくつか示す。

- ・天候不順の影響などにより、衣料品分野の売上は厳しいまま推移いたしました
- ・売上原価の増加積極的な研究開発活動を行い、研究費が増加したことなどにより、営業利益 583 億 47 百万円 (前年同期比 26.3% 減)、経常利益 609 億 22 百万円 (前年同期比 25.8% 減) となりました

テストデータの 343 文の業績要因文に対して人手にて極性を付与して正解データを作成し、本手法の精度、再現率を求めた。結果を表 3 に示す。比較手法として、人手によって極性付与された手がかり表現のみを使用し、業績要因文に含まれている手がかり表現の極性に従って、業績要因文の極性を付与する手法の精度、再現率を表 4 に示す。なお、もし「推移する」のように極性を付与できない手がかり表現のみが含まれていた場合は、極性を付与しない。

³<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

表 4: 業績要因文への極性付与の評価結果 (比較手法)

	positive	negative
精度 (%)	93.0 (120/129)	89.8 (62/69)
再現率 (%)	54.5 (120/220)	50.4 (62/123)

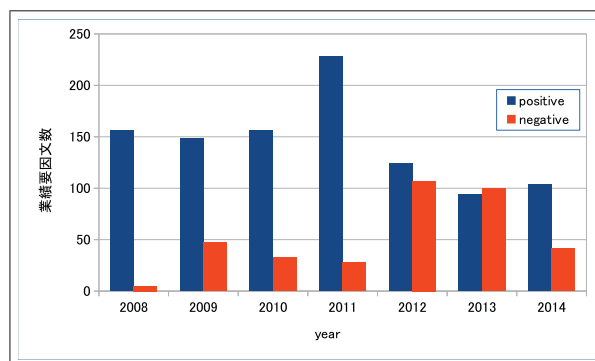


図 1: 「太陽電池」を含む業績要因数の推移

5 考察

本手法による業績要因文への極性付与手法の精度はポジティブで 86.7%、ネガティブで 90.0% であり、高い精度を達成した。また、再現率はポジティブで 95.4%、ネガティブで 73.9% であり、比較的良好な結果を得ることができた。一方で、人手によって極性付与された手がかり表現のみを使用する比較手法では再現率がポジティブで 54.5%、ネガティブで 50.4% であることから、本手法は比較手法と比べて、それほど精度を落とすことなく再現率が大きく向上していることが分かる。例えば、本手法では以下の業績要因文に対してポジティブの極性を付与することができたが、比較手法では極性を付与することができなかった。

セラミックス事業は、アジアを中心としたガラスエンジニアリング分野と国内の環境エネルギー分野で需要が旺盛であったため、概ね堅調に推移しました

本手法は、上記のような「推移しました」のような極性を付与できない手がかり表現を含む業績要因文に対しても極性を付与することができるため、比較手法に比べて再現率を向上することができた。

決算短信には発表年月日が記載されているため、あるキーワードを含む業績要因文が記述されている決算短信の発表日を得ることで、そのキーワードを含むポジティブな業績要因文数、ネガティブな業績要因文数の時系列推移を得ることができる。例として、図 1 に「太陽電池」を含む業績要因文数の時系列推移を示す。図 1 より、2011 年において「太陽電池」を含むポジティブ

な業績要因文数は最も多いが、2012年になると大幅に減少し、逆にネガティブな業績要因文数が大幅に増えていることが分かる。2012年の「太陽電池」を含むネガティブな業績要因文には、以下のような記述がある。

多結晶シリコンは、太陽電池パネルの供給過剰とパソコンの販売不振等を背景にした半導体ウエハーの在庫調整に伴う、販売数量の減少及び販売価格の下落により大幅な減収減益となりました

以上のように、極性付与した業績要因文は、企業における製品や事業の現状の把握から、製品や事業の動向の時系列推移とその要因など、様々な適用が可能である。

6 関連研究

酒井らは、経済新聞記事の企業の業績発表記事から抽出した業績要因文に対して極性を自動的に付与する手法を提案している [5]。酒井らの手法は、業績発表記事に対してSVMで高精度で極性付与できることが前提となっており、酒井らの手法をそのまま決算短信に適用するのであれば、決算短信に対して高精度で極性を付与するための、大量の学習用データを人手で用意する必要がある。高村らは2つの単語から成る表現に対して隠れ変数モデルを用い、機械学習を用いて構成語の属性をクラスタという形で抽出して確率モデルを構築し、複数語表現の感情極性を分類する手法を提案している [7]。それに対して、本手法では業績要因文を抽出するために有効な手がかり表現（「が好調でした」等）に極性を付与し、その手がかり表現を拡張した“拡張手がかり表現”を使用することで学習データを自動生成し、深層学習により業績要因文への極性付与を行った。

Kajiらは、評価文に含まれる評価表現（名詞＋格助詞＋形容詞）に対して極性を決定する際に、人手で作成した手がかり表現リストやパターン、規則を使用して評価文を抽出して極性を付与しておき、評価表現が好評文に出現する頻度、不評文に出現する頻度を使用することで評価表現の極性を決定する手法を提案している [2]。この手法では、同一の評価表現が不評文、好評文に3回以上、出現する必要があるが、業績要因文は多くの名詞や動詞で構成されるため、同一の業績要因文が出現することはまずない。そのため、本手法では業績要因文に含まれる手がかり表現のみに着目して、業績要因文への極性付与を行った。

7 まとめ

本稿では、決算短信より抽出された業績要因文に対して深層学習を用いて極性（「ポジティブ」、「ネガティブ」）を自動的に付与する手法を提案した。具体的には、業績要因文を抽出するための手がかりとなる表現にくつつかの文節を追加することで、“拡張手がかり表現”と定義する文節列を獲得した。そして、“拡張手がかり表現”を使用することで極性付与のための学習データを自動的に生成し、生成された学習データを使用して深層学習を行い、“拡張手がかり表現”を含まない業績要因文に極性を付与した。評価の結果、業績要因への極性付与手法の精度はポジティブで86.7%、ネガティブで90.0%、再現率はポジティブで95.4%、ネガティブで73.9%であり、良好な精度、再現率を達成した。

ブ)を自動的に付与する手法を提案した。具体的には、業績要因文を抽出するための手がかりとなる表現にくつつかの文節を追加することで、“拡張手がかり表現”と定義する文節列を獲得した。そして、“拡張手がかり表現”を使用することで極性付与のための学習データを自動的に生成し、生成された学習データを使用して深層学習を行い、“拡張手がかり表現”を含まない業績要因文に極性を付与した。評価の結果、業績要因への極性付与手法の精度はポジティブで86.7%、ネガティブで90.0%、再現率はポジティブで95.4%、ネガティブで73.9%であり、良好な精度、再現率を達成した。

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, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3309–3315 (2011).
- [2] Kaji, N. and Kitsuregawa, M.: Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents, *Proceedings of the EMNLP-CoNLL 2007*, pp. 1075–1083 (2007).
- [3] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834–1842 (2002).
- [4] Sakai, H. and Masuyama, S.: Cause Information Extraction from Financial Articles Concerning Business Performance, *IEICE Trans. Information and Systems*, Vol. E91-D, No. 4, pp. 959–968 (2008).
- [5] Sakai, H. and Masuyama, S.: Assigning Polarity to Causal Information in Financial Articles on Business Performance of Companies, *IEICE Trans. Information and Systems*, Vol. E92-D, No. 12, pp. 2341–2350 (2009).
- [6] 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. J98-D, No. 5, pp. 172–182 (2015).
- [7] 高村大也, 乾孝司, 奥村学: 隠れ変数モデルによる複数語表現の感情極性分類, 情報処理学会論文誌, Vol. 47, No. 11, pp. 3021–3031 (2006).