

# 株式・債券・為替ごとの金融極性辞書の自動構築

## Automatic Construction of Financial Sentiment Lexicons on Stock, Bond and Exchange

伊藤 友助<sup>1</sup> 酒井 浩之<sup>1</sup> 北島 良三<sup>2</sup>

末廣 徹<sup>3</sup> 稲垣 真太郎<sup>3</sup> 木村 柚里<sup>3</sup>

Tomosuke Ito<sup>1</sup>, Hiroyuki Sakai<sup>1</sup>, Ryoza Kitajima<sup>2</sup>

Toru Suehiro<sup>3</sup>, Shintaro Inagaki<sup>3</sup>, Yuri Kimura<sup>3</sup>

<sup>1</sup> 成蹊大学

<sup>1</sup> Seikei University

<sup>2</sup> 東京工芸大学

<sup>2</sup> Tokyo Polytechnic University

<sup>3</sup> みずほ証券株式会社

<sup>3</sup> Mizuho Securities Co., Ltd.

**Abstract:** In this paper, we propose a method of retrieving expressions and constructing financial sentiment lexicons on stock, bond and exchange. Retrieved phrases are distinctive words in each financial market and extensions of these words. For example, the sentiment of the word "increase" in the financial sentiment lexicons cannot be determined by merely looking at the word itself. However, we can determine sentiment by considering the relation between preceding and subsequent words, such as "increased consumption" and "increase in costs". In addition, it is necessary that the expression such as "cause inflation" determines appropriate sentiment on stock, bond and exchange. We acquire automatically these expressions, and we construct appropriate financial sentiment lexicons by stock, bond and exchange.

## 1. はじめに

投資家は様々なデータを踏まえて投資判断を行うが、そのデータは膨大で多岐にわたるため、近年は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。その一例が「利上げ」のような金融に関連する語を収録した金融極性辞書である。

金融極性辞書について、これまでに株式に特化した「ネガティブ・ポジティブ」付与の研究が行われている。しかし、その極性辞書を債券などの他の金融市場に当てはめると極性が異なる場合が多い。これは同じ言葉でも金融市場によって極性が異なる場合があるためであり、例えば、「物価上昇」は株式市場においてはポジティブと判断するが、債券市場においてはネガティブと判断する必要がある。また、既存の金融極性辞書には単語（あるいは複合語）が収録されており、例えば、「増加」という語について、一つの極性が付与されている。しかし、「消費が増加」や「コストが増加」のように、語の前後関係まで広

く捉えてみると、必ずしもその語に付与されている極性が正しいとは限らないことがわかる。

そこで本研究では、金融テキストを用いて、株式・債券・為替の金融市場ごとに表現にまで拡張した文字列での金融極性辞書を自動構築するための手法を提案する。本研究によって、辞書に収録するのに適した表現を獲得し、それらに極性を付与することで、各金融市場に特化した金融極性辞書の自動構築を実現する。

本研究では、金融極性辞書構築のためのデータとして、日経 QUICK 月次調査と日本経済新聞の記事を使用した。日経 QUICK 月次調査は、センチメントなどの定量的なデータが示されており、各金融市場のテキストデータを有するため、市場ごとの分析が可能である。

これまでに金融極性辞書についての研究は行われている[1][2][3][4]が、株式に特化した辞書が多く、辞書に収録されているのは単語（あるいは複合語）単位である。本研究では、株式・債券・為替の金融市

場ごとに語の前後関係まで拡張した表現を自動的に獲得する点異なる。また、同一の表現でも金融市場によっては極性が異なる場合が考えられる。そのため、極性を付与するための基準となる情報を、例えば既存研究[3][4]のように株価のような1つの指標にすることはできず、各金融市場において適する指標を基準として、各表現に対して極性付与を行う必要がある。

## 2. 提案手法の概要

本研究では、日経 QUICK 月次調査と日本経済新聞の記事を用いて、金融市場ごとに表現にまで拡張した文字列での金融極性辞書の自動構築を行う手法を提案する。金融極性辞書には各金融市場において特徴的な語を表現にまで拡張した文字列を格納する。提案手法の概要を以下と図1に示す。

**Step 1:** 日経 QUICK 月次調査と日本経済新聞の記事を用いて、株式・債券・為替の金融市場ごとに金融極性辞書に収録するのに適した語の候補（以降、「金融市場特徴語候補」と定義）を抽出する。

**Step 2:** 酒井らの手法[5]を用いて獲得した手がかり表現と抽出した金融市場特徴語候補をもとに、金融市場ごとに特徴的な語を表現にまで拡張した文字列（以降、「金融市場特徴表現」と定義）を獲得する。

**Step 3:** 獲得した金融市場特徴表現に極性の付与を行い、金融極性辞書を構築する。

## 3. 金融市場特徴語候補の抽出

### 3.1 金融市場特徴語候補抽出の概要

**Step 1:** 日経 QUICK 月次調査に記載されている各金融市場担当者からのコメント文（以降、「コメント文」と定義）を抽出し、コメント文と日本経済新聞記事をもとに Doc2vec のモデルを金融市場ごとに生成する。

**Step 2:** 生成した Doc2vec のモデルを用いて、金融市場ごとに類似した記事を日本経済新聞記事から抽出し、金融市場ごとにコメント文と合わせた文章データ（以降、「拡張文章データ」と定義）を作成する。

**Step 3:** 酒井らの手法[5]を用いて、拡張文章データから金融市場ごとに共通頻出表現・手がかり表現を獲得する。

**Step 4:** 拡張文書データをもとに、Word2vec のモデルを金融市場ごとに生成する。

**Step 5:** 生成した Word2vec のモデルに獲得した共通頻出表現・手がかり表現を入力し、モデルから出力された語を金融市場特徴語候補として金融市場ごとに抽出する。

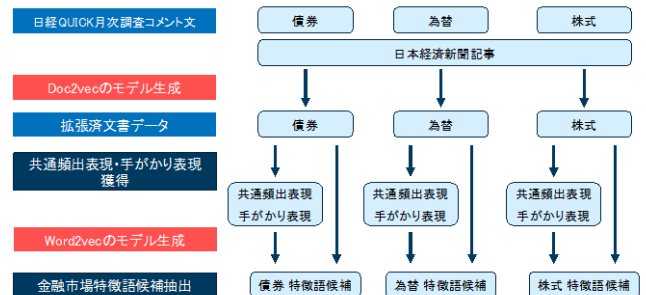


図1 金融市場特徴語候補抽出の概要

### 3.2 Doc2vec のモデル生成と類似記事抽出

日経 QUICK 月次調査に記載のコメント文を金融市場ごとに抽出する。しかし、抽出したデータが金融市場特徴語を獲得するデータ量として不足しているため、それらのテキストデータをもとに Doc2vec を用いて、日本経済新聞記事より金融市場ごとに類似する記事を抽出する。各金融市場においてそれぞれ Doc2vec のモデルを生成し、日経 QUICK 月次調査に記載のコメント文集合との類似度を求め、上位 50,000 記事に関連記事としてコメント文と合わせて拡張文書データとする。

### 3.3 共通頻出表現・手がかり表現獲得

酒井らの手法[5]を用いて、金融市場特徴語候補を抽出するための手がかりとなる表現を金融市場ごとの拡張文書データからそれぞれ自動的に獲得する。共通頻出表現・手がかり表現を獲得する手法について以下に述べる。なお、具体的な獲得手法については、文献[5]を参照されたい。

**Step 1:** 少数の手がかり表現（今回は「が好調」「が不振」を用いる）を人手で与え、それに係る節を取得する。

**Step 2:** 取得した節の集合から、その中で共通して頻繁に出現する表現（「売り上げ」など）を共通頻出表現として抽出する。

**Step 3:** 共通頻出表現に係る節を取得し、その中から新たな手がかり表現を抽出する。

**Step 4:** 獲得した手がかり表現から、それに係る節を取得する。

**Step 5:** Step 2 から Step 4 を新たな手がかり表現と共通頻出表現が獲得されなくなる、もしくは、予め定めた回数まで繰り返す。

獲得した共通頻出表現・手がかり表現は形態素解析し、名詞と接頭詞、動詞の場合のみ結合した複合語とする。各金融市場における共通頻出表現・手がかり表現の例を以下に示す。

・株式

圧迫する, 輸出関連株, 下値不安, 好調, 収益率, 不安材料, 修正, 経常赤字, 落ち込む, 維持する, 一巡し, 業績回復, 押し下げ, 借り入れ, 弱まる

・債券

引き上げる, 消費増税, 高齢者, 財政悪化, 売却, 先行し, 縮小する, インフレ, 債務残高, 新興国, 先行き不透明感, 急減し, 招き, 浮上する, 税込

・為替

後押しする, ユーロ高, 期待感, 政情不安, 後退, 投げ売り, 再燃し, リスク回避姿勢, 利上げ観測, 中央銀行, 増大する, 下支えし, 良好, 財政支出

おける金融市場特徴語候補の例を以下に示す。

・株式

原価低減, 半導体需要, 自社株買い, 可処分所得, 危機感, 軽自動車, 買い戻し, 失速懸念, 薄らぐ, アジア向け輸出, 潜在成長率, 停滞感, くすぶり

・債券

ドル通貨オプション取引, 銀行間市場, ベア, 夢, 中国経済, 投資妙味, 上向く, デフレ圧力, 急伸, 円高基調, 低成長, 買い進め, 下げ止まる, 顕著

・為替

原油依存度, 日中韓, 代わる, 需給懸念, 鎮める, 中国人民銀行, 衰退, 米輸出入物価指数, 叫ばれ, 柔軟姿勢, 取り崩す, 実効為替レート, 手詰まり感

### 3.4 Word2vec のモデル生成と金融市場特徴語候補抽出

拡張文書データから金融市場特徴語候補を自動で取得するため、Word2vec を活用する。

Word2vec は単語の分散表現（ベクトル）を得る手法である。単語をベクトルで表現することで、単語間のコサイン類似度を計算することができ、意味のよく似た単語（類似度が高い単語）を抽出することができる。Word2vec のモデル生成時の学習データ（金融市場ごとの拡張文書データ）によって、同じ入力語であっても出力される類似度が高い単語が異なるため、金融市場における特徴的な語が出力されることを期待する。

なお、学習データは形態素解析したのち、連続する品詞や接頭詞の場合は結合し、複合語として作成した。金融市場ごとに先の学習データによって生成したモデルを用いて、前節で獲得した共通頻出表現・手がかり表現を入力とし、類似度が 0.5 以上の語を金融市場特徴語候補として抽出する。各金融市場に

### 3.5 金融市場特徴語候補の抽出結果

本手法により獲得した共通頻出表現・手がかり表現の数の詳細を表 1 に示す。また、抽出した各金融市場における金融市場特徴語候補の数の詳細を表 2 に示す。なお、「単独」はその金融市場にのみ存在する語の数を示しており、「総数」はその金融市場において得た語を合計した数を示している。

表 1 各金融市場の共通頻出表現・手がかり表現獲得数

金融市場	単独	総数
株式	170	1,230
債券	175	1,219
為替	72	1,096

表 2 各金融市場の金融市場特徴語候補の数

金融市場	単独	総数
株式	1,560	6,978
債券	1,494	7,009
為替	880	6,147

## 4. 金融市場特徴表現の獲得

酒井らの手法[5]を用いて獲得した手がかり表現と抽出した金融市場特徴語候補をもとに、金融市場ごとに金融市場特徴表現を自動獲得する。表現獲得の概要を以下に示す。図 2 は株式市場における金融市場特徴表現の獲得の例を図示したものである。

#### 4.1 金融市場特徴表現の獲得の概要

**Step 1:** 獲得した共通頻出表現と手がかり表現の中から、手がかり表現のみを金融市場ごとに抽出する。（「が好調」「が不振」など）

**Step 2:** 金融市場特徴語候補と抽出した手がかり表現を結合し、金融市場ごとに表現にまで拡張した文字列を生成する。

**Step 3:** 生成した文字列が表現として適しているか判定を行い、文字列を金融市場特徴表現として獲得する。

・株式市場の例

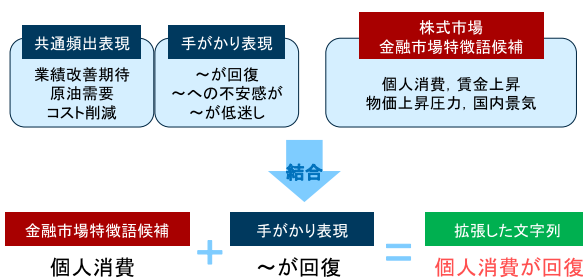


図2 株式市場における金融市場特徴表現獲得の例

#### 4.2 表現にまで拡張した文字列の判定

本手法では、生成した文字列が表現として適しているか判定するために、拡張文書データを用いた。具体的には、金融市場ごとに金融市場特徴語候補と手がかり表現を結合して文字列を生成したのち、その文字列が金融市場ごとの拡張文書データに含まれているかの確認を行う。含まれていた場合、その文字列を金融市場特徴表現として獲得する。

#### 4.3 表現獲得の結果

本手法によって獲得した金融市場特徴表現の一例を以下に示す。また、各金融市場における金融市場特徴表現獲得数の詳細を表3に示す。なお、表3には各金融市場において、拡張文書データに含まれていた回数ごとの表現の数、ならびに表現獲得数の総数を示している。

・株式

景況感指数が改善した、賃金上昇が続き、先行き不安が強い、市場が回復、輸出関連株が上昇し、個人消費への懸念が、景気減速懸念が強まった、需給が悪化して、物価上昇率が低迷、リスクが低く

・債券

米経済が回復し、均衡が拡大し、成長率が下がる、不透明感が強まる、赤字幅が縮小、政権が継続、信用リスクへの警戒感から、相場が伸び悩む、低金利政策が続く、言及がなかった、購入が必要

・為替

リスク回避姿勢が再燃、ドル高基調が高まる、財政赤字が続き、マイナス幅が縮小している、負債の多い、金利が低下、政策への期待感から、生産が増加、貿易赤字が続いて、警戒感が後退して

表3 各金融市場の金融市場特徴表現獲得数

金融市場	1回	>2回	>4回	総数
株式	53,148	20,522	11,343	86,759
債券	52,573	18,948	10,022	84,178
為替	43,647	16,176	8,751	70,386

#### 4.4 金融市場特徴表現の評価

金融市場特徴表現についての評価は、金融市場ごとに獲得した表現のうち、拡張文書データに3回以上含まれている表現の中からランダムにそれぞれ100個抽出し、それらを人手によって正解を判定して行った。正解判定の基準は、ある表現について、その表現を獲得した金融市場において考えたときに、極性を付与することが可能かどうかで判断した。評価結果の精度を表4に示す。

表4 評価結果

金融市場	精度(%)
株式	88.0
債券	92.0
為替	80.0

### 5. 金融市場特徴表現への極性付与

本手法により獲得した金融市場特徴表現に対して、対象としている市場に適した極性の付与を行う。しかし獲得した表現の数は多く、市場ごとに人手にて適切な極性を付与することは多大な労力を要する。そこで最小限の労力で極性付与のための学習データを自動的に生成し、深層学習により極性を付与することを試みた。極性付与の概要を以下に示す。

## 5.1 金融市場特徴表現への極性付与の概要

**Step 1:** 獲得した共通頻出表現と手がかり表現の中から、手がかり表現のみ（「が好調」「が不振」など）を金融市場ごとに抽出し、その中から手がかり表現単体で考えたときに、ネガティブまたはポジティブが明らかなものに対してその極性を人手で付与する。

**Step 2:** 金融市場ごとに拡張文書データ（類似する日本経済新聞記事上位 50,000 記事のみ）を文単位に分割し、文中に 4 章で獲得した金融市場特徴表現が含まれている文のみ抽出する。（以降、「テストデータ」と定義）

**Step 3:** テストデータの中から Step 1 で極性を付与した手がかり表現が含まれている文に対してその極性を付与し、学習データを作成する

**Step 4:** LSTM を用いた学習モデルを生成し、モデルをもとにテストデータに極性の付与を行う。

**Step 5:** テストデータに含まれる金融市場特徴表現に対して、テストデータである文に付与された極性をもとに極性の付与を行う。

## 5.2 手がかり表現への極性付与

3 章において、酒井らの手法[5]を用いて、金融市場特徴語候補を抽出するための手がかりとなる表現を金融市場ごとの拡張文書データからそれぞれ自動的に獲得を行った。その中から「が好調」「が不振」といった手がかり表現のみを金融市場ごとに抽出する。

金融市場ごとに抽出した手がかり表現の中から手がかり表現単体で考えた時に、その金融市場においてネガティブまたはポジティブが明らかなものに対して極性を人手で付与する。以下に、株式市場における手がかり表現への極性付与の例を図 3 に挙げる。

株式市場 手がかり表現	
～が回復	.... ポジティブ
～が低迷し	.... ネガティブ
～が増加	.... 不明

図 3 手がかり表現への極性付与の例

例のように、「～が回復」は、前にくる金融市場特徴語候補に関わらずポジティブであると判断できるため、手がかり表現に極性の付与を行う。同様に「～が低迷し」も明らかにネガティブの表現であると判

断できる。しかし、「～が増加」については、前に出現する金融市場特徴語候補によって極性が変化するという。例えば、「リスクが増加」では株式市場においてネガティブと判断するが、「売り上げが増加」ではポジティブと判断する必要がある。このような手がかり表現に対しては、手がかり表現単体で考えたときに極性が明らかではないため、極性の付与は行わない。

## 5.3 学習データ・テストデータの作成

4 章にて獲得した金融市場特徴表現に極性を付与するための学習データを作成する。

まず、3 章にて作成した金融市場ごとの拡張文書データの中から、金融市場ごとに類似する日本経済新聞記事 50,000 記事のみを抽出し、それらを文単位に分割する。記事単位であるデータを文単位に分割する理由としては、記事は全ての文において一つの極性が付与されているわけではなく、ネガティブ、ポジティブの文が混在して構成されており、記事への極性はそれらの文を総合的に判断して付与される。本研究の目的は、金融市場特徴表現への極性付与であり、文単位に極性を付与することによって、記事単位に極性を付与した場合よりも表現に対して正確な極性が付与できると考えるからである。

文単位に分割したのち、4 章で獲得した金融市場特徴表現が文中に含まれている文のみを抽出し、これらをテストデータとする。さらに、テストデータの中から人手で極性付与した手がかり表現が含まれている文にはその極性を付与し、複数の手がかり表現が含まれる場合には最後に出現する手がかり表現の極性を文に付与し、極性が付与された文をまとめて学習データとする。

株式市場における学習データの例を以下に示す。なお、先頭の 1, 0 は、学習データに付与された極性を示し、1 がポジティブ、0 がネガティブを表す。

1 000622-0220/382 家計の所得環境の緩やかな改善などを映し個人消費が堅調に推移するうえ、企業の設備投資も増加に転じる。  
0 100929-1327/165 デフレが進行する中では、企業や家計の名目所得も大きく減少することから、税収が減少し、財政再建もおぼつかないだろう。

図 4 に本研究にて用いた LSTM+多層パーセプトロン (MLP) による学習モデルを示す。

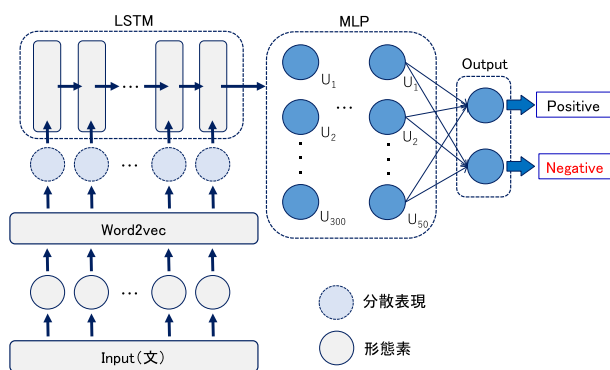


図4 LSTM+多層パーセプトロン (MLP)

学習は以下の手順で行う。

**Step 1:** 学習データの文を形態素解析し、各語の分散表現 (300次元) を LSTM に入力する。

**Step 2:** 文頭から文末までのベクトルを多層パーセプトロン (MLP) の入力層 (300ユニット) に入力する。

**Step 3:** MLP の出力層は2ユニットとなっており、それぞれ、positive, negative の尤度を出力する。

MLP の中間層は9層 (ユニット数 200 が3層, ユニット数 100 が3層, ユニット数 50 が3層) である。エポック数は株式市場においては20, 債券市場, 為替市場においては50とし, 活性化関数はランプ関数を使用した。また, 分散表現を得るために word2vec を使用し, そのデータには日本経済新聞記事10年分を用いた。

本学習モデルは文の形態素の分散表現が文頭から文末まで推移していく過程が組み込まれたモデルであり, 学習データにおける文の文脈を考慮した学習を行うことができる。

#### 5.4 金融市場特徴表現への極性付与

作成した学習データを用いて上記の学習モデルにより, まずはテストデータに極性を付与する。テストデータである文のネガティブ, ポジティブそれぞれの尤度と, その文に含まれる金融市場特徴表現を獲得する。なお, 尤度が高い方を優勢とし, その文に付与する極性とする。

次に, テストデータ各文のネガティブ, ポジティブの尤度とその文に含まれる金融市場特徴表現をもとに, 金融市場特徴表現に対して極性の付与を行う。

ある文に含まれている金融市場特徴表現について, その文におけるネガティブ, ポジティブの尤度がその表現の尤度と等しいこととする。また, 同じ金融市場特徴表現でも含まれる文によって尤度は異なる

ため, 同じ表現のネガティブ, ポジティブの尤度の平均値をその金融市場特徴表現の尤度とする。これは一つの表現においてネガティブ, ポジティブの尤度を一意に決めるためである。

金融市場特徴表現の極性は, その表現のネガティブ, ポジティブの尤度が高い方を優勢とし, その極性とする。さらに, ネガティブ, ポジティブの尤度のどちらかが0.7以上の表現のみを抽出する。

各金融市場における金融市場特徴表現への極性付与数を表5に示す。また, 株式市場における金融市場特徴表現への極性付与の例を表6に示す。

表5 各金融市場における金融市場特徴表現への極性付与数

金融市場	総数
株式	59,027
債券	61,590
為替	63,580

表6 株式市場における金融市場特徴表現への極性付与の例

極性	金融市場特徴表現	negative 尤度	positive 尤度
positive	復興需要への期待から	0.227	0.772
positive	消費心理が回復	0.015	0.984
positive	中間決算が好調だった	0.011	0.988
negative	警戒感が残る	0.723	0.276
negative	倒産が増加	0.829	0.171
negative	マイナス材料が増えて	0.957	0.044
negative	住宅バブルが後退して	0.784	0.216

## 6. 評価

本手法についての評価は, 極性の付与を行った金融市場特徴表現のうち, テストデータ内で出現する頻度が高いものから順に表現の先頭2文字が本評価のために抽出する表現と重複しない表現を100個抽出し, それらを人手によって正解を判定して行った。評価対象の表現について, 先頭2文字が他の評価対象の表現と重複しない表現にする理由としては, 表現の前半部分である金融市場特徴語候補が偏ることなく, 多様な表現を評価するためである。なお, 正解判定の基準はある表現について, その表現を獲得した金融市場において考えた時に, 付与されている極性が正しいかどうかで判断した。株式市場, および, 債券市場における評価結果を表7に示す。

表7 評価結果

金融市場	精度(%)
株式	75.0
債券	42.0

## 7. 考察

金融市場特徴表現の獲得について、表4より、本手法による精度は株式 88.0%、債券 92.0%、為替 80.0%と高い精度を達成することができた。これは、金融市場特徴語候補と獲得した手がかり表現を結合し、その文字列が各金融市場に特化した拡張文書データに3回以上含まれている場合に金融市場特徴表現として評価対象としたためであると考えられる。獲得した表現は、実際に投資判断に役立てられている金融テキストデータに含まれている表現であるため、金融極性辞書に収録するのに適している語が多い。

しかし、表現が含まれていた回数を増やすことで精度が上がるとは言い切れず、例えば、株式の拡張文書データ内において5回出現する「状況がいつ」という表現が挙げられる。この表現は拡張文書内において、「状況がいつ好転するか不明である」などのような文の一部であると考えられる。このように、文書内において多く出現する表現であったとしても、極性を付与することができない表現が抽出されていることもあり、極性を付与するのに適していない表現の除去する手法が必要である。

次に獲得した金融市場特徴表現における極性付与について、本手法では尤度が0.7以上の表現のみを抽出して極性を付与したが、尤度の閾値を設定しない場合は、株式市場において精度が46.0%という結果となり低かった。精度が低い理由としては、その金融市場特徴表現（本研究では、金融市場特徴語候補と手がかり表現を結合したもの）単体で考えた時に極性の判断ができないものが存在していることが考えられる。例えば、表8のような表現である。

表8 金融市場特徴表現への極性付与の例（尤度が低い場合）

極性	金融市場特徴表現	negative 尤度	positive 尤度
positive	拡大が続く	0.468	0.532
positive	地合いが続く	0.480	0.520
positive	注目度が高い	0.499	0.501
negative	動きが強まった	0.583	0.417
negative	傾向が続く	0.631	0.369
negative	ムードが強い	0.607	0.393

これらの表現は、その表現だけでは極性の判断が難しく、表現のさらに前にくる単語や表現によって極性を決めることができると考える。

評価において不正解と判定された表現について注目をしてみると、ネガティブとポジティブの尤度がそれぞれ0.5に近いものが多いことがわかった。どちらかの尤度が高い表現は本研究において極性がはっきりしていると言えるため、どちらかの尤度が0.7以上の表現のみを抽出し、再度評価を行った。株式市場において、評価に用いた金融市場特徴表現の例を表9に示す。

表9 金融市場特徴表現への極性付与の例（どちらかの尤度が高い場合）

極性	金融市場特徴表現	negative 尤度	positive 尤度
positive	効果が期待できる	0.016	0.984
positive	景気が回復	0.100	0.898
positive	材料が必要	0.266	0.734
negative	業績が悪化	0.929	0.071
negative	低迷が続く	0.845	0.155
negative	信用力の低い	0.722	0.278

これらの表現を用いて再度同様の方法で評価を行った結果が表7に示されている。表7より、株式市場において、精度が75.0%と比較的良好な結果を得ることができた。このことから、金融極性辞書には、表現のネガティブ、ポジティブのどちらかの尤度が0.7以上である金融市場特徴表現を格納することがふさわしいといえる。

また、表7より、債券市場における精度が42.0%と低かった。これは、債券市場は株式市場と比較して、極性の付与が複雑であることが理由だと考える。債券の場合は、一般的に景気が悪くなると安全資産であるため人気が上昇する。そのため表現の対象が景気であれば、ネガティブな表現でも極性をポジティブと判定しなくてはならない。例えば「景気が低迷」はポジティブと判定する必要がある。しかし、表現の対象が債券であれば、ネガティブな表現をそのままネガティブと判断する必要がある。例えば、「債券価格が低迷」という表現があげられる。そのような極性付与をするためには、学習データの作成時に手がかり表現の極性ではなく、その手がかり表現が対象としているものが「景気」なのか「債券」なのかを判定し、学習データを作成する必要があるが、本手法では学習データに含まれている手がかり表現の極性に従って極性を付与しているため、対応

できていなかった。

なお、為替市場の表現への極性付与ならびに金融極性辞書の構築も本手法と同様の方法で作成を行うことができるが、対象が「円」である場合と「ドル」である場合で付与すべき極性が変化する。そのため、債券の場合と同様に学習データ作成時に対象についても考慮して作成する必要があると考える。

## 8. まとめ

本研究では、金融テキストを用いて、株式・債券・為替の金融市場ごとに表現にまで拡張した文字列での金融極性辞書を自動構築するための手法を提案した。

具体的には、日経 QUICK 月次調査と日本経済新聞の記事を用いて、はじめに株式・債券・為替の金融市場ごとに金融市場特徴語候補を抽出する。そして、獲得した手がかり表現と抽出した金融市場特徴語候補をもとに、金融市場ごとに金融市場特徴表現を獲得した。獲得した表現における評価の結果、精度は株式 88.0%、債券 92.0%、為替 80.0%と高い精度で表現を獲得することができた。

その後、獲得した金融市場特徴表現に極性の付与を行った。極性付与の結果を検証するためにテストデータ内で出現頻度が高く、かつ、評価対象の表現間で前半部分の重複がなく、ネガティブ、ポジティブのどちらかの尤度が 0.7 以上の表現を抽出した結果、株式市場において、精度が 75.0%と比較的良好な結果を得ることができた。

今後の課題としては、金融極性辞書に格納されている表現について、更に精度を向上させることである。特に債券市場、為替市場において表現の極性付与のための学習データ作成について別の手法も検討し、より最適な極性を付与することで、精度の向上を目指す必要がある。また、本研究では、各金融市場において金融市場特徴表現を格納している。しかし、単語または複合語ベースで最適な極性を付与することができる語も存在しており、それらに極性を付与し、表現とともに本研究の金融極性辞書に格納することでより実用性の高い極性辞書になることが期待できる。

## 参考文献

- [1] Ito T., Sakaji H., Tsubouchi K., Izumi K., Yamashita T. (2018) Text-Visualizing Neural Network Model: Understanding Online Financial Textual Data. In: Phung D., Tseng V., Webb G., Ho B., Ganji M., Rashidi L. (eds)

Advances in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science, Springer, vol 10939, pp 247-259.

- [2] 伊藤諒, 和泉潔, 須田真太郎: ネットワークの表現学習による金融専門極性辞書の構築, 2017 年度人工知能学会全国大会(第 31 回), 2017.
- [3] 五島圭一, 高橋大志: 株式価格情報を用いた金融極性辞書の作成, 自然言語処理, Vol.24, No.5, pp547-577, 2017.
- [4] 関和広, 柴本昌彦: 銘柄固有の金融極性辞書の構築, 第 18 回金融情報学研究会, 2017.
- [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.