

予想材料に基づく為替の動向情報分析を目的とした ニュース記事分類手法

Classifying News Articles with Factors for Foreign Exchange Forecast

岩崎有基^{1*} 松下光範¹
Yuki Iwasaki¹ Mitsunori Matsushita¹

¹ 関西大学大学院
¹ Kansai University

Abstract: The purpose of this study is to support reviewing on their own decision for foreign exchange trends. Foreign exchange trends can be predicted from economic and political events. This is a decision based on experience, and it enables improvement of accuracy by verifying the validity of own analysis. However, as the factors of foreign exchange fluctuations are diverse, it is difficult to verify the adequacy of foreign exchange trends. To solve this problem, this paper proposes a method to classify news articles based on investor analysis factor in order to grasp the relevance of news articles and foreign exchange trends.

1 はじめに

外国為替取引(以下、為替取引と記す)や株式売買は資産運用の一つとして広まっている。しかし、投資における資産運用には損失を被るリスクがあり、投資家はリスク軽減のために為替や株の動向を分析し、分析内容を根拠として通貨売買の意思決定を行う。

為替取引において、為替動向を分析する手法はテクニカル分析とファンダメンタル分析の2通りに大別される。テクニカル分析は為替レートの数値情報に基づいて、為替相場の変動を予測する手法であり、ファンダメンタル分析は経済や政治のニュース記事といったテキスト情報から為替相場に対する影響を見積もる手法である。これらの分析手法を統合的に用いることで投資家は為替動向の分析を行う。

ファンダメンタル分析を行う場合経済や政治に関する出来事から為替動向を予測するが、この行為は経験による判断が大きい。岩崎らの研究において熟達投資家は投資市場における外的要因を考慮しリスクテイク量(自身の資産のうちどの程度の投資を行うか)を調整していたが、非熟達投資家はどのような場合にも同じリスクテイク量をとっていた[6]。このように投資家は熟達度によって経済や政治に関する出来事といった外的要因を考慮する度合いが異なる。

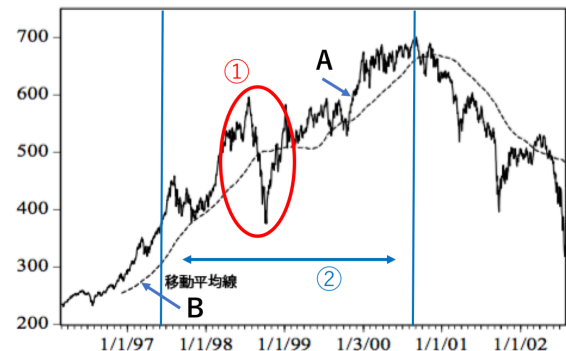


図 1: 移動平均線 (200) の例 [3] より

為替の動向分析において投資家は自身の分析精度を向上させるために分析から導き出した動向予測と実際に市場が動いた結果を照らし合わせ、自身の分析の妥当性を検証する必要がある。しかし、為替相場における要因は多様で非熟達投資家はどの要因が為替相場に影響を与えたのかを判断することが難しい。そこで、本研究は為替動向に対する自らの判断についての振り返りを支援することで、非熟達投資家の分析精度の向上を目的とする。熟達投資家と同じニュース記事の情報整理を非熟達投資家も行えるようにすることで分析精度の向上が期待される。そこで、本稿ではニュース記事と為替の関連性把握のために、ニュース記事を投資家の分類指標に基づいて分類する手法を提案する。

*連絡先: 関西大学大学院総合情報学研究所知識情報学専攻
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: k732146@kansai-u.ac.jp

2 為替における動向分析手法

本章ではテクニカル分析とファンダメンタル分析の詳しい説明と本研究における着眼点としてファンダメンタル分析を選択した理由を述べる。

2.1 テクニカル分析

テクニカル分析は過去の為替レートの値動きを用いることで未来の相場を分析する手法であり、主に数値情報を用いる。代表的な分析手法として移動平均線(図1-B)と呼ばれるものがある。移動平均線は分析困難な価格変動を平滑化し方向性(e.g., 円安への動きや円高への動き)の傾向を視覚化することができる。例えば、図1-Aを分析する際に図1-①の部分に着目すると下落方向の動きに見える。しかし、図1-②の期間を見ると下落ではなく上昇の一部である。このとき、図1-①内の点線(図1-B)に着目すると、右肩下がりでないことから下落ではなく、上昇の一部だとわかる。このように、テクニカル分析は次節で述べるファンダメンタル分析と比べ、価格変動のみに着目することで分析が容易である。

2.2 ファンダメンタル分析

ファンダメンタル分析はマクロ経済(e.g., インフレ, 原油価格, 失業率), 各業界特有の情報(e.g., コンペ, 需要と供給, 技術進化), 各企業特有の情報(e.g., 成長率, 業績, 訴訟事件)などの情報を用いて市場の動きを予想する分析方法である。ファンダメンタル分析では主にニュース記事のテキスト情報が用いられ、ニュース記事に記述されている経済や政治に関する出来事は為替の相場変化の要因となる。例えば、2009年のリーマン・ショックではリーマン・ブラザー・ホールディングスの倒産を契機に米ドル/円が円高に推移したように経済や政治に関する出来事は直接的に為替に影響を及ぼす。

ファンダメンタル分析はテクニカル分析に比べてニュース記事を俯瞰的に解釈する必要がある。そのため分析することが難しく、数値化することができず明確な指標が存在しない。また、ニュース記事の解釈は人により異なる。例えば、2016年の米大統領選挙ではトランプ氏の勝利により円高に推移すると予想する場合とトランプ氏の政策によりアメリカが成長するため円安に推移するといった2通りの考え方が存在した。また、ファンダメンタル分析ではマクロ経済学やミクロ経済学などの多方面の分野に精通していなければ理解が難しい。

表 1: 予想材料 (和泉ら [5] の研究より)

予想材料	元となる生データ
景気	GDP, NAPM, GNP 指数
物価	消費者物価指数, 卸売物価指数
金利	公定歩合, 長期金利
マネーサプライ	マネーサプライ
貿易収支	貿易収支
雇用	失業率, 非農業就労者数
個人消費	小売売上, 個人所得
介入	介入
要人発言	中央銀行総裁等の発言
マルク	ドルマルク, 円マルクのレート
石油	石油価格
政治	政情, 国際的な事件
株	株価
債券	債券価格
短期トレンド 1	先週の変動値
短期トレンド 2	変動値の変動値
長期トレンド	5 週間の変動

2.3 本研究の立ち位置

テクニカル分析やファンダメンタル分析は一方の分析方法のみ使用して為替動向の分析を行うことができる。しかし、テクニカル分析は統計的な手法のため過去にない動きなどは予想できない。また、ファンダメンタル分析は突発的な事象による為替動向の予測はできるが、テクニカル分析よりも大雑把にしか予測ができない。そのため、これらを統合的に用いることが為替動向の分析では重要である。

ファンダメンタル分析はテクニカル分析と比べ、明確な指標がなく、俯瞰的に解釈する必要があるため分析が難しい。そこで、本研究ではファンダメンタル分析に焦点を当てて支援を行うことで、非熟達投資家の分析の精度の向上を図る。

3 関連研究

外国為替取引市場において、投資家は大量のニュース記事などを用いて分析する。このような背景からニュース記事を為替動向の予測のために用いている研究は多い。Deshらは短い期間において為替の動向をリアルタイムのニュースヘッドラインから予測するテキストマイニング技術を提案した[4]。ニュースヘッドラインは最も重要なニュースの要約が記述されている。そのためニュースヘッドラインから予測を行うことが可能となっている。例えば、ニュースヘッドラインには“BOJ SEEN KEY CALL RATE STEADY ON THURSDAY”と

表 2: スクレイピングしたデータセット

head	date	body
米 F R B、2 回のレポで...	2008 年 01 月 1 日 10:11 JST	[ニューヨーク 3 1 日 ロイター] 米連邦準備理事会 (F R B) 傘下の...
前タイ首相派の国民の力党...	2008 年 01 月 1 日 10:20 JST	[バンコク 3 1 日 ロイター] 2 3 日のタイ下院選挙で第一党と...
北朝鮮、期限までに核計...	2008 年 01 月 1 日 10:25 JST	[ワシントン 3 1 日 ロイター] ケーシー米 국무省副報道官は 3 1 日、...

表 3: 整形したデータセット

head	date	body
米 F R B、2 回のレポで...	2008/01/1	米連邦準備理事会 (F R B) 傘下の...
前タイ首相派の国民の力党...	2008/01/1	2 3 日のタイ下院選挙で第一党と...
北朝鮮、期限までに核計...	2008/01/1	ケーシー米 국무省副報道官は 3 1 日、...

記述されているとする。「停滞している (STEADY)」、「上昇した (UP)」などの情報からシステムは為替動向を予測を行う。リアルタイムの取引に対してはこの技術は有用だが、トレード後の検証時に振り返る際、どんな要因によって変動したかを詳しく知ることができず、自身の経験として蓄積されない。そこで、本稿ではニュース記事本文から為替変動の要因を抽出することでニュース記事を分類し、どの要因が為替変動に影響を与えたのかをユーザが確認できるようにする。

株価の分析においてもニュース記事が主な判断材料として使用されるため、ニュース記事を使用した予測や分析といった研究も存在する。馬場らはトピックモデルの一つである LDA (Latent Dirichlet Allocation) [1] を拡張したヘテロトピックモデルを用いて株に関するニュースの分析を行う手法を提案している [7]。ニュース記事は投資の判断材料となるような企業の情報を示していることが多いことから、株価への影響を与えている可能性が高い。そこで、馬場らは株価だけでなく、ニュース記事も用いることで統合的な分析環境の構築を行った。記事の分類の指標として用いているものは、LDA によりニュース記事群から取得した潜在的なトピックである。ヘテロトピックモデルは LDA で取得したトピックに対し株価情報を考慮してそのトピックが企業に影響を与えるかの判定を行う。この研究では潜在的なトピックを LDA でニュース記事から取得することで分類を行っていたが、必ずしもそれが投資家が分析する際に行う分類方法とは限らない。本稿では熟達投資家の分類指標に基づいて分類を行うことで分析を行うための基盤を構築する。

福元らは単語の共起関係を RF(Random Forest)[2] や Jubatsu¹ を利用してニュース記事のカテゴリ分類を行い、複数の学習アルゴリズムについて検証している

¹<http://jubat.us/ja/>

[8]。ニュース記事を分類する際にニュース記事をベクトルデータとして扱うが、次元数が膨大になってしまう問題がある。そこで福元らはシソーラスの単語の意味属性を用いて共起頻度による共起行列を生成することで次元数の削減を図った。それらを RF や Jubatsu の学習データとして用い、識別率の比較を行っている。この研究では次元削除のために共起関係を用いている。本研究では gensim² を使用し、ベクトル化に用いる単語を出現回数から選定し、次元削除を行うことで RF の学習データとして用いる。

4 デザイン指針

本章では、ニュース記事を分類するにあたり考慮した要因を述べる。

4.1 予想材料

本研究ではニュース記事を分類する指標として和泉らの研究内で明らかとなっている予想材料を用いる [5]。この予想材料は人工市場のマルチエージェントが考慮する内容を決定するために、熟達投資家にインタビューを行い、為替相場が動く要因となるものを分類したものである。この予想材料を分類指標として用いることによって熟達投資家と同じニュース記事の整理を行うことができ、ユーザの分析の効率を上げることができると考えた。

和泉らの研究で用いられている予想材料は 17 種類 (表 1) ある。しかし、予想材料の一つであるマルクはドイツの旧通貨であり、1999 年に廃止され 2018 年 11 月現在用いられていない。また、短期トレンドや長期

²<https://radimrehurek.com/gensim/> (2018/11 確認)

トレンドはテクニカル分析に用いるもので、為替レートといった価格が元となる生データである。そこで本研究では 17 種類のうち 13 種類(景気, 物価, 金利, マネーサプライ, 貿易収支, 雇用, 個人消費, 介入, 要人発言, 石油, 政治, 株, 債券)を用いてニュース記事分類を行う。

4.2 使用するデータ

2.2 節で述べたように、ファンダメンタル分析において経済や政治の出来事を為替動向変化の要因として使用している。個人投資家の経済や政治の出来事を知る情報源はニュース記事のため、本研究においてニュース記事を学習データとして用いている。ニュース記事分類モデル作成に用いる学習データはロイターのトップニュース 2008 年 1 月 1 日から 2009 年 12 月 31 日の 2 年分を用いる。ニュース記事は Web 上からスクレイピングを行い抽出している。取得したニュース記事の総数は 23,969 件である。表 2 にスクレイピングしてきたデータの一部を示す。1 列目にはニュース記事のタイトル, 2 列目にはニュース記事が公開された時間(以下, 公開日), 3 列目には実際に記述された記事の全文(以下, 本文)が入力されている。本文の文頭にはどの地域で最初に発行されたかを示す内容が記述されており, ニュース記事分類の要素として不必要なため, 前処理として除いている。また, 公開日の書式と何時何分に発行されたかという詳細な時間まで必要ないため除いた。これらの条件のもと, 整形したデータを用いたものが表 3 である。本研究では, このように収集し整形したニュース記事を RF の学習データとして用いる。

5 提案手法

和泉らの予想材料の手がかりではニュース記事内に存在するそれぞれの予想材料の特徴を全て包括できているとは言えない。本研究では予想材料ごとにそれぞれの特徴語を収録している辞書を作成し, それを元にニュース記事にラベル付けを行うことで学習データを作成した。

5.1 辞書の作成

ニュース記事を予想材料ごとに分類するために予想材料ごとの特徴語をまとめた辞書を作成した。辞書は予想材料の元となる生データ(表 1)を参考に, 日経シ

ソーラス³と Yahoo! ファイナンス⁴の経済指標カレンダーから用語を抽出した。日経シソーラスは金融や政治の用語が階層構造になっている。例えば, 金利の子には貸出金利, 金利据え置き, FF 金利などが含まれる。それらを予想材料(金利)の特徴語として辞書を作成し, 景気, 物価, マネーサプライなども同様に行い作成した。また, Yahoo!ファイナンスの経済指標カレンダーからは, 経済指標と呼ばれる政府が発表する GDP などの統計をそれぞれの特徴に入力した。しかし, この方法での辞書作成では, 政治政策の新しい用語(e.g., リーマンショック, オバマケア)や要人は任期によって変化することから, 政治と要人発言についての特徴語が日経シソーラスから抽出できなかった。そこで, 政治, 要人発言を除いた 11 種類の特徴語群を用いて記事を分類した。分類方法は, 辞書内にある予想材料の特徴語の単語の記事に含まれていれば, その予想材料の記事としている。例えば, ニュース記事中に「FF 金利」という言葉があれば, その記事は金利の予想材料に属すると考え分類を行う。11 種類のどの特徴にも属しなかった記事に政治や要人発言の予想材料が含まれると仮定し, 政治, 要人発言と思われるニュース記事を人手で分類した。人手で分類する際に, 政策に関する記述, 「~を述べた」や「~との見解を示した」といった発言と思われる記述があればそれらの条件を元にそれぞれ政治, 要人発言としている。

以上の 13 種類の記事群から, 頻出している単語の抽出を行い記事群の特徴語として辞書の拡張を行った。この拡張により「日経シソーラス」や「Yahoo!ファイナンス」から作成した用語だけでなく, 実際にニュース記事内で用いられている単語を特徴語とすることができる。異なる予想材料の特徴語が重複しないために, 特徴語が重複した場合は TF-IDF 値により, 値の大きいもののみを用いた。TF-IDF は TF(Term Frequency)と IDF(Inverse Document Frequency)を合わせたものである。TF はすべての文章にわたって頻出している単語を重要語とし, IDF はすべての文書に横断的に存在する単語の重要度は低いとしている。この TF と IDF を組み合わせることで, すべての文章に横断している単語の重要度は低く, 他の文章中に存在しない単語ほど重要度を高くできる。TF-IDF 値は scikit-learn⁵の tfidfVectorizer を用いて算出した。表 4 に以上の工程から作成された辞書の一部を示す。

³[urlhttp://t21.nikkei.co.jp/public/help/contract/price/20/thesaurus/](http://t21.nikkei.co.jp/public/help/contract/price/20/thesaurus/), (2018/11 確認)

⁴[urlhttps://info.finance.yahoo.co.jp/fx/marketcalendar/](https://info.finance.yahoo.co.jp/fx/marketcalendar/), (2018/11 確認)

⁵<http://scikit-learn.org/>, (2018/11 確認)

表 4: 辞書のラベルと辞書データの例

ラベル	特徴語の例
景気	GDP, GNP
物価	消費者物価指数, 卸売物価指数
金利	公定歩合, 長期金利
マネーサプライ	資金移動, マネーサプライ
貿易収支	貿易収支, 貿易黒字
雇用	失業率, 非農業就労者数
個人消費	小売売上, 個人所得
介入	市場介入, ドル買い介入
要人発言	オバマ大統領, 日銀総裁
石油	原油在庫, 原油価格
政治	大統領選挙, 辞任表明
株	日経平均, ダウ
債券	社債, 個人向け国債

ニュース記事

米連邦公開市場委員会で**フェデラルファンド金利**の誘導目標を...FRB当局者は依然、経済の下振れリスクを懸念しているが、一方で原油や食料の値上がりを受けた**インフレ**圧力にも目配りしている

予想材料	景気	物価	金利	マネーサプライ	貿易収支	雇用	個人消費	介入	石油	株	債券	政治	要人発言
ラベル	1	0	1	0	0	0	0	0	0	0	0	0	0

図 2: ニュース記事へのラベル付与の例

5.2 辞書を用いたラベル付け

5.1 節で作成した辞書を用いて学習用のラベルを作成した。本研究では図 2 にニュース記事本文に対してラベル付けを行った例を示す。左から順番に、景気、物価、金利、マネーサプライ、貿易収支、雇用、個人消費、介入、石油、株、債券、政治、要人発言である。辞書内に存在する予想材料の特徴語がニュース記事本文に存在しているかを判定しラベル付けをしている。ニュース記事本文中に辞書内の単語が存在している場合は 1 を、存在していない場合は 0 を付与した。上記の方法でラベル付けしたところ、図 2 のようになった。この図からわかるようにニュース記事は予想材料単体のラベル(シングルラベル)だけでなく、複合してラベルが付与される(マルチラベル)ことがある。このことからニュース記事には複数の予想材料が含まれていることがわかる。そのため、それぞれの記事をマルチラベルからシングルラベルへの変換を行った。あるニュース記事が内包している予想材料のラベルの数だけニュース記事を複製し予想材料のラベルを単一に割り当てた。例えば、図 2 の場合だと景気と金利の予想材料を内包している記事なので 2 つの同じニュース記事が生成され、それぞれに景気や金利とラベル付けされる。このように 2008 年 1 月 1 日から 2009 年 12 月 31 日までの 23,968 件の記事に対してラベル付け処理を行った。

6 ニュース記事ごとの分類実験

本章では、5.2 節で作成したデータセットを元に scikit-learn の RF を学習させ、ニュース記事の分類を行い、正解率、適合率、再現率、F 値によるニュース記事分類モデルの評価を述べる。

6.1 実験の流れ

5.2 節で作成したデータセットを用い教師あり学習である RF を学習させ、学習後のモデルを使用しニュース記事の分類を行った。データセットの文は MeCab を用いて一般名詞や固有名詞を抽出しその他の動詞や形容詞、副詞可能、非自立、代名詞、助詞類接続、加えて 12 月などの数値情報などを除いて Bag-of-Words の形式にしている。また、「オバマ/大統領」といった連続して名詞が出現している場合「オバマ大統領」と名詞同士を結合している。その後これらの名詞情報からニュース記事をベクトル化することで、それぞれのニュース記事ごとの特徴量としている。その際、gensim を使用し、ベクトル作成に用いる単語はすべての記事にわたり 20 回以上出現し、5 割以上の記事に出現していないものとしている。これらを学習データ 36,819 件 (6 割)、テストデータ 24,546 件 (4 割) に分割して使用した。

RF はシード値を 0 に固定し、作成する決定木の数は 10 としている。この条件の元、上記の特徴量を RF に学習させ、テストデータの正解と予測データの比較を行った。比較を行った際の評価指標である正解率、適合率、再現率、F 値は以下としている。

$$\text{正解率} = \frac{\text{正解したデータ数}}{\text{予測した全体の記事数}} \quad (1)$$

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (2)$$

$$\text{recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (3)$$

$$\text{F-measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

上記の評価指標により算出された正解率、適合率⁶、再現率⁷、F 値⁸を表 5 に示す。学習データセットの分け方による正解率の違いを図るために 5 分割交差検定も行った。その平均値は 0.122 となった。これらの数値は小数点第四位を四捨五入を行い算出している。

⁶http://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html

⁷http://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html

⁸http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

表 5: 記事ごとの分類評価

正解率	適合率	再現率	F 値
0.113	0.425	0.113	0.179

表 7: 文ごとの分類評価

正解率	適合率	再現率	F 値
0.769	0.878	0.767	0.819

表 6: 辞書のラベルと辞書データの例

記事番号	文番号	本文	ラベル
0	0	米連邦準備理事...	00000000000000
0	1	内訳は 4 日物が...	00000000000000
0	2	受け入れ担保は...	00100000000000
0	3	2 日物は米国債...	00000000000000
0	4	応札額は 4 日物...	00010000000000

6.2 考察

ニュース記事ごとのラベル付けで RF に学習させたところ正解率, 適合率, 再現率, F 値はどれも低いものとなった。これは 5 分割交差検定の結果から学習に用いたデータセットの組合せの問題ではないことがわかる。データデータセット作成のために本文をベクトル化した, その際に他の特徴も含まれていたため, それらが影響し正解率などが低下したと考えられる。例えば, 図 2 をラベルの数だけ複製し, それぞれに景気や金利といったラベルを付与していたが, この場合ニュース記事には景気と金利のベクトルが本文に存在している。そのため学習を行う際にノイズとなり正解率やその他の評価指標の低下につながったと考えられる。そこで, ニュース記事ごとのラベル付けではなく, ニュース記事を文ごとに分割しラベル付けを行うことで余分なベクトルデータを削減する方法での学習を行う。

7 文ごとの分類実験

本章では, 6.2 節の考察を元にデータセットの再作成を行い, ニュース記事分類モデルを評価した。

7.1 実験の流れ

5.2 節ではニュース記事をラベルの数だけ複製し, それぞれに景気や金利といったラベルを付与していた。この場合それぞれの予想材料の特徴がベクトルデータに含まれてしまうため, ニュース記事を分割し, 一文ごとにラベルを付与することで特徴ベクトルの削減を行った。あるニュース記事を分割しラベル付けを行った例を表 6 に示す。6.1 節のように MeCab を使用し, Bag-of-Words の形式でニュース記事文のベクトル化を行っている。その際 6.1 節と同じく, 一般名詞や固有

名詞を残し, その他の動詞や形容詞, 副詞可能, 非自立, 代名詞, 助詞類接続, 12 月などの数値情報などを除いて作成したデータセットを使用し RF の学習を行った。データの総数は 214,094 件あり, それを学習データ 128,454 件 (6 割), テストデータ件 85,640 (4 割) に分割し用いている。正解率, 適合率, 再現率, F 値を表 7 に示す。また, 5 分割交差検定を行ったところ平均値は 0.757 となった。

7.2 考察

7.1 節から予想材料を元にニュース記事を分類する場合, ニュース記事ごとに分類するよりも有用である。ニュース記事を文ごとに分割して分類した方がそれぞれの予想材料のラベルに適した特徴ベクトルをそれぞれの文が保持しているため, 正解率, 適合率, 再現率, F 値が向上した。また, ニュース記事を文により分類することが可能になれば, ニュース記事をセグメンテーション化できるようになる。そうすることで, ニュース記事のどの部分が金利の予想材料なのか, 景気の予想材料なのかをユーザに提示することが可能となり, ユーザはより詳しくニュース記事に記述されている為替動向の変化要因を知ることができる。以上の正解率, 適合率, 再現率, F 値とセグメンテーション化することができる有用性からニュース記事ごとに分類するよりも, ニュース記事の一文ごとに分類する方が有用である。

8 今後の展望

今後は本研究で作成したニュース記事分類モデルを一機能として実装したインタフェースの開発を行う。図 3 に理想のインタフェースを示す。このインタフェースはユーザの為替動向の分析を支援することを目的として作成する。

図 3-A には為替レートのグラフが表示され, 図 3-B にはニュース記事のタイトルが表示されている。左図に表示されているグラフは 1 日ごとの高値, 安値, 始値, 終値が表示されており, 折れ線グラフより詳細な情報を確認することができる。右図にはニュース記事のタイトルが表示されているが, このタイトルをクリックすることでその記事のタイトルがハイライトされ, 本文を表示することができる。また, その選択したニュース記事がグラフ上のどの位置の記事がを提示してくれる。記事本文が表示された際には本文下部に 3 種類のボタ

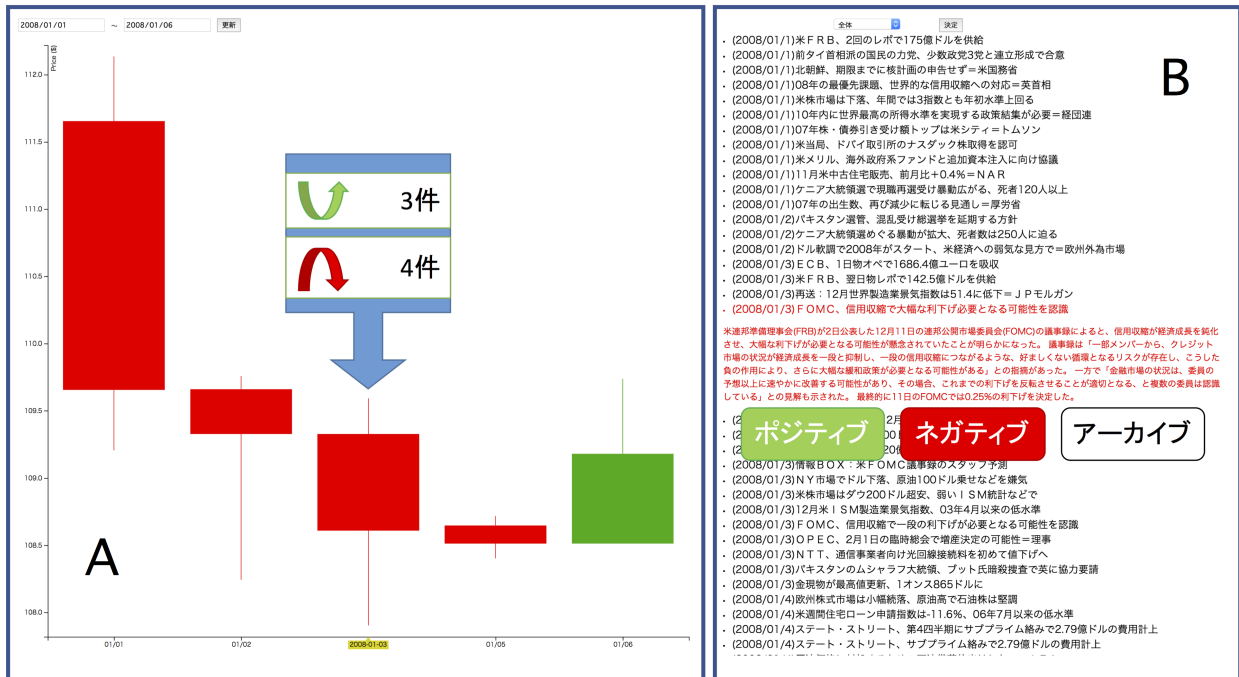


図 3: 理想のシステム図

ンが表示され、ユーザはそのニュース記事が為替動向の上昇要因か、下落要因か、それとも関係がないかをのタグを付けることができる。そうすることで、振り返りの際に図 3-A 上に対応するグラフの日付とどのようにユーザ自身がニュース記事を分類したかをシステム側から提示され、自身の思考の流れを思い返すことができる。図 3-A 上部にはプルダウン式で日付が選択でき、表示したい期間を指定することができる。この期間指定の機能はグラフの表示期間だけでなく、ニュース記事の期間も指定され、指定された期間外のニュース記事は表示されなくなる。図 3-B の上部に表示されているプルダウン部分は何の予想材料を表示したいかを選択する際に使用でき、例えば景気を選択した場合、景気予想材料を含んでいる記事が提示される。

このようなインタフェースを実装することで、実際にユーザは分析の効率を向上させられるかについての検証を行う予定である。

9 おわりに

本研究では為替動向に対する自らの判断についての振り返りの支援を目的に、本稿ではニュース記事と為替の関連性把握のために、ニュース記事を熟達投資家の分析指標に基づいて分類する手法を提案する。ニュース記事ごとのラベル付けを行ったデータセットとニュース記事 1 文ごとのラベル付けを行ったデータセットで

RF を学習させ、ニュース記事分類モデルを作成した。それぞれのニュース記事分類モデルを正解率、適合率、再現率、F 値で評価したところ、ニュース記事ごとのラベル付けよりもニュース記事を文ごとに分割してラベル付けしたデータセットで学習したニュース記事分類モデルの方が有用だと示唆された。

謝辞

本研究の遂行にあたり、文部科学省科学研究費 (課題番号:15H02780) の助成を受けた。記して謝意を表す。

参考文献

- [1] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation, *Journal of machine Learning research*, Vol. 45, No. 1, pp. 993–1022 (2003)
- [2] Breiman, L.: Random forests, *Machine learning*, Vol. 45, No. 1, pp. 5–32 (2001)
- [3] Griffioen, G. A. W.: Technical analysis in financial markets, Doi: <http://dx.doi.org/10.2139/ssrn.566882>, University of Amsterdam (2003)

- [4] Peramunetilleke, D. and Wong, R. K.: Currency exchange rate forecasting from news headlines, *Australian Computer Science Communications*, Vol. 24, No. 2, pp. 131–139 (2002)
- [5] 和泉潔, 植田一博, 中西晶洋: 人工市場モデルによる外国為替ディーラーの学習行動の分析, 情報処理学会研究報告,1996-ICS-106, Vol. 106, No. 105, pp. 91–98 (1996)
- [6] 岩崎雄斗, 和泉潔, 伊藤祐輔, 植田一博: 投資家のリスク態度と熟達度ならびに市場動向が投資行動に与える影響, 日本認知科学学会論文誌, Vol. 22, No. 3, pp. 389–408 (2015)
- [7] 馬場慧, 馬強: 株価とニュースの統合分析のためのヘテロトピックモデル, データ工学と情報マネジメントに関するフォーラム (2017)
- [8] 福元伸也, 淵田孝康: 単語の共起関係に基づく機械学習による文書分類, 情報処理学会研究報告, Vol. 2014-DBS-160, No. 28, pp. 1–5 (2014)