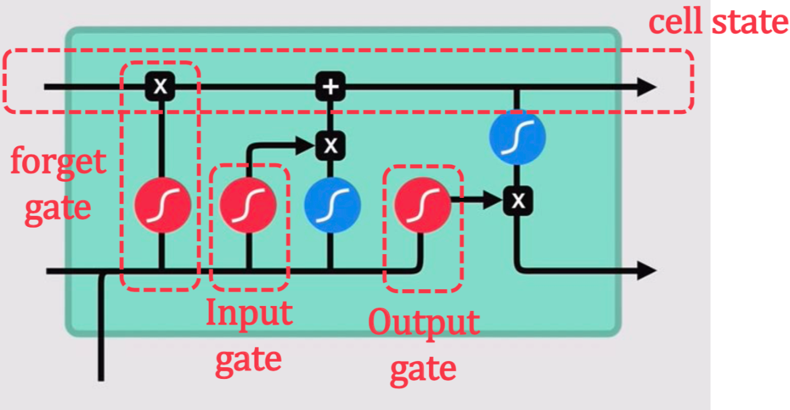
|  |
| --- |
| **FX deep learning + NLP(Trump tweet)** |
|  |
| **要約**  LSTMは様々な時系列予想問題及び時間分類問題にうまく適用されていて、株、FXの予想に適している。しかし、過去のデータからLSMTで機械学習するのには金融ニュースを校了することができないという問題点がある。今回、Natural Langage Processing(NLP)のSentimentalを使ってトランプ大統領のTweetが及ぼすFX市場の影響を考慮して正確なトレンドをPython nltkで学習させるのと、過去のデータから数分、数時間後のトレンドをLSMTで学習させるのをあわせた新しい指標を使って市場のトレンドを予想する。 | | |

1. **プロジェクトの動機と背景**

今日、世界の金融市場は現代経済に多忙な影響を及ぼしており、様々な市場で　、毎日数秒ドルの金融商品が取引されている。外国為替市場(FX)もこれに例外ではない。すべてのトランザクションは、時間とともに変動する通貨の為替レートを形成する。毎日の取引で高額な取引を考えると、これらのレートの変化を予測することができれば、莫大な利益の可能性がある。

LSTM(Long Short−Term Memory)

　Long Short-Term Memory networkは通常「LSTM」と呼ばれており、長期的な依存関係を学習することができる特別なRNNである。Hochreiter & Schmidhuber(1997)によって発表され、後続の研究者たちがたくさんの論文を書き、現在も広く使われている技術です。short-term memoryとは短期記憶のことであり、短期記憶を長期にわたって活用することを可能にしたのが、LSTMの重要な成果である。RNNを実現するために考案されたため、前の情報をうまく処理しようとする目的は一緒であるがRNNとは中身の処理が違う。



Cell Stateは相対的な情報を終わりまで転送する。学習結果はここを通る。Forget gateは前の入力と今回の入力が重要かそうでないかを判断してCell Stateにどのくらい伝達するかどうかをきめる。Input gateはCell stateをどのくらいアップデートするのかを決める。Output gateは次の入力に情報を伝達する。[1][2][3]

Foreign Exchange(FX)

　正式名称は外国為替証拠金取引という。**異なる通貨の売買**のことで、円やドル、円やウォン、ドルやユーロ、ポンドなどの通貨そのものを買ったり売ったりする取引である。通貨のレートは常に変動していて、FXはこの変動するレートを使って利益を出します。例えば、１円＝１０ウォンの時に１円を買い、一ヶ月後１円＝１２ウォンになっていたら２ウォンの利益になります。もし、１００万円買っていたら２００万ウォンの利益になります。しかし、利益を上げるためには**通貨がいつ安くいつ高いのかという相場の流れを知り**、決済のタイミングを知る必要がある。海外情勢に左右されるFXのレートは、チャートを見て過去の値動きの傾向を把握しながら取引をしていくのが一般的である。[4]



1. **既存研究内容**

１分毎にトレンドが上下するのかを予想するテクニカル分析にいくつかの深層学習を適用するもので、使用モデルはConvolutional Neural NetworkとLSTM Neural Networkであった。[1]

1. **問題定義**

既存研究内容で説明した研究ではテクニカル分析しか使われていない。つまり過去のFX取引データだけをみてトレンドを判断している。しかし、実際にトレンドの上下を判断するには金融ニュースも考慮してトレンド予想しなくてはならない。金融ニュース記事には、金融商品の価格またはレートに影響を与える重要な情報が含まれていることが多いが、今回は情報は考慮されていない。つまり、ファンダメンタル分析(経済的指標)は使われていないのが今回の問題である。

1. **研究目標**

本研究では問題定義を考慮してテクニカル分析(レート解析)に加え、ファンダメンタル分析(経済的なニュース、要人の発言)を考慮して自然言語学習をさせる。ほかのFX予測問題ではNLPを考慮してるものはなく、今回始めてLSTMとNLPをあわせた新しいモデルを作っていくことにする。

1. **先行研究**

テクニカル分析

テクニカル分析とは、**過去の相場の動きから、将来の値動きを予想する分析方法**である。過去の値動きをグラフにしたチャートなどを使用する場合が多々ある。指標にはローソク足、サポートラインとレジスタンスライン、フィボナッチリトレースメント、移動平均、RSI、MACD、ストキャスティクス、ボリンジャーバンド、ダウ理論などがある。

移動平均

**ある一定期間の価格（通常は終始）を平均で結んだ線**

基本は短期線、中期線、長期線の３種類に分けることができる。一般的に短期線は５日、６日、１２日。中期線は２０日、２５日、５０日。長期線は１００日、１４４日、２００日、２３３日が用いられる。**このような平均線を見ることでトレンドを把握することができる**。[4]

既存のDeep Learning LSTMモデル機械学習

LSTMモデルに合うようにデータを前処理する。LSTMの連続する値は移動平均期間に合わせて決める。論文では短期移動平均線5日間を使用している。まずトレーニングデータを5日間で一つにまとめる。各5日間の移動平均を求める。ターゲットにするものは、前の移動平均から後ろの移動平均を引いたとき、値が0よりも大きかったら前の期間より値が上昇したとみなし+1の答えを、値が０よりも小さかったら前の期間より値が下降したとみなし-１の答えを与える。使用したパラメータはActivation Function(relu, softmax)、Opimizer(Adam)、Loss function(Categorical CrossEntropy)である。[1]

ファンダメンタル分析

ファンダメンタル分析とは**為替に関する基礎的な情報をもとに相場を分析する**ことである。経済指標や金利、その他通貨や国に関する様々な要因が分析材料となっている。要因には経済指標の他にも重要人物の発言も影響される。例えばFRB(Federal Reserve Board）議長、ECB(European Central Bank)総裁、日本銀行総裁の発言が注目される。その中でもSNSを通してDonald TrumpのTweetがFX市場に影響を与えている。2018年7月23日のロイターの調査でトランプとFXの関係性に関する記事があった。[4]



自然言語の前処理

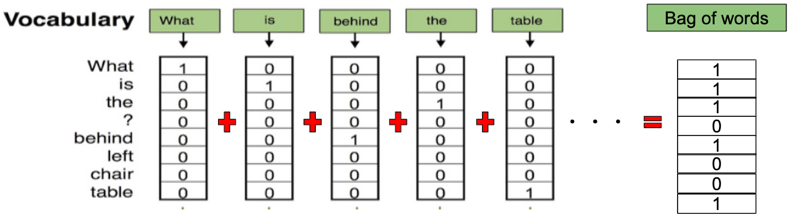
前処理の大まかな流れとして、クリーニング処理→文章の単語分割→単語の正規化→ストップワードの除去と大まかに流れる。まずクリーニング処理では、~~<h1>~~FX news usd jpy~~</h1>~~みたいにテキスト中のノイズを除去する作業をする。文章の単語分割では、原文「私は今日海に行きます」という文章を「私/は/今日/海/に/行き/ます」のように分割させる。単語の正規化では、Appleをappleみたいに文字種の統一や大文字小文字変換をする。ストップワード除去では、タスクを解くために役立たないと考えられる単語を除去する。[7]

Bag of Words

Bag of Wordsというのは自然言語処理において自然言語（人間が日常で使用している言語）で記述されたデータ、つまり文をベクトルで表現する方法のことである。

Bag of Wordsは次の３ステップで作ることができる。[8]

1. 数値変換
2. one hot vector
3. 足し合わせる



1. **問題解決方法**

先行研究を通して問題定義で掲げている自然言語処理をあわせたLSTMのFXトレンド予想を実現するために以下の解決方法を定義することにした。

1.短期、中期、長期移動平均をLSTMの形式にデータの前処理をする。

2.WebからFXに関するニュースを抽出して自然言語処理をする。

3.(2)からFXに関する重要単語を取得する。

4.Donald TrumpのTweetをAPIで取得し、その中で(2)で取得したFX用語が含まれているTweetを選択し、自然言語処理をする。Tweetが肯定的か否定的かを調べ、機械学習のターゲットにする。[5][6]

5.実際にDeep LearningのLSTMモデルを使用して機械学習をする。

6.一般的なDeep Learning LSTMモデルと今回のモデルとは比較する。

短期、中期、長期移動平均をLSTMの形式にデータの前処理をするでは、今回短期移動平均を使い実際のデータのshapeは(Total FX data, 5 days, columns number)という形になる。WebからFXに関するニュースを抽出して自然言語処理をするでは、googleでUSD/JPYに関する記事をweb scrapingして取得し、自然言語前処理をした。つぎにBag of wordsを使いそのニュースでよく使われている名詞単語リストを作成する。Trump tweetを前処理し、名詞単語リストに含まれている単語があるtweetだけを選び感情分析をする。肯定的なら+1、否定的なら-1とターゲットにする。これをLSTM機械学習モデルを使用して学習させ、既存のモデルと比較をする。

データの分け方

Twitter APIだけでなくKaggleのデータも使う

Twitter API

(Train data) 8 : 2 (Test data)

Date : 2019/07/14 ~ 2019/11/15

Train data : 2019/07/14 ~ 2019/10/25

Test data: 2019/10/26 ~ 2019/11/15

Kaggle

(Train data) 8 : 2 (Test data)

Date : 2015/07/17 ~ 2016/11/11

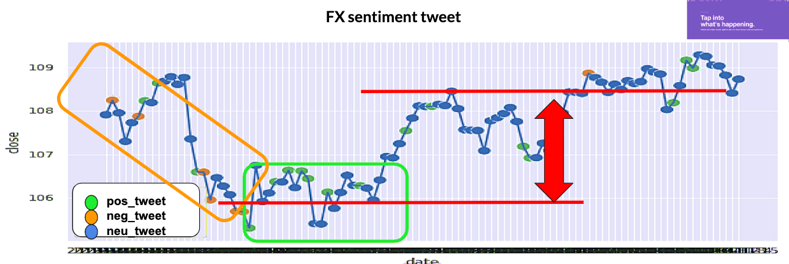
Train data : 2015/07/17 ~ 2015/08/05

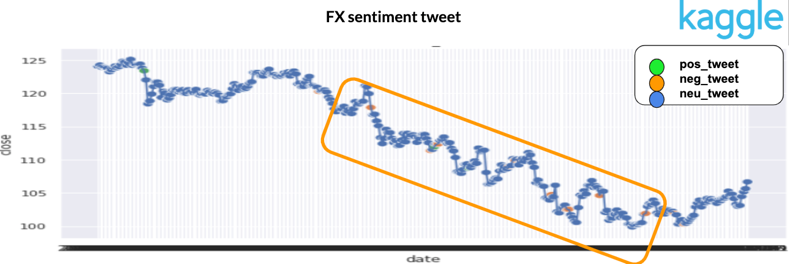
Test data: 2015/08/06 ~ 2016/11/11

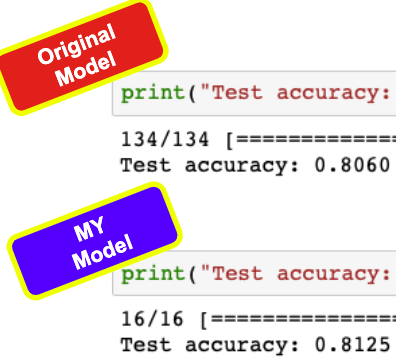
1. **結果・評価**

　緑が肯定的Tweet、オレンジが否定的Tweet、青が中立のTweetである。実際に肯定的、否定的で分け可視化してみた結果、否定的なTweetが集まっているところではトレンドが下がる傾向が見られ、肯定的なTweetが集まれば後にトレンドが上昇する力が働くことがわかりました。これを通してトランプ大統領のTweetが少なからずFXに影響があるということがわかりました。

LSTM機械学習もさせ、既存モデルと今回のモデルの比較をした結果、少しだけですが今回のモデルの方が正確性が高かったことがわかりました。







1. **Discussion**

　まずなによりも惜しかったのはTwitter APIの制限で2019/07/15~2019/11/15までという十分なTweet数を取得することができなかったことです。だが今回の研究はLSTMとNLPをあわせたFXの研究が言及された先行研究の事例がなかったため、新しい指標を使って研究しました。その結果LSTMでFXトレンド予想とNLP学習によって経済的な指標も少しだけ考慮することができた。Twitterの制限問題が解決すればもっと良い研究結果が出るはずである。

1. **研究の重要性**

　今回の研究でDonald TrumpのTweetがFXのレートに関わっていることがわかった。そして過去の動きのパターンと経済的な指標を考慮してFXトレンドを深く知ることができた。この研究が最終的に利益を得ることができるための参考になれば良いと思ってます。

1. **今後の展望**

　Web scraping技術を駆使して経済的な指標のデータを大量に集める。また、他の研究者がこの研究を参考にしてバージョンアップした研究をすすめることを願っています。

1. **参考文献**

[1]COMPARING CONVOLUTIONAL AND LSTM NEURAL NETWORKS FOR INTRADAY FOREX DIRECTION PREDICTION(2018)

[2]Forecasting of Forex Time Series Data Based on Deep Learning(2019)

[3]いまさら聞けないLSTM(2017)

[4]FX外国為替証拠金取引デモトレードの実行と分析(2011)

[5] Characterizing Social Relations Via NLP-based Sentiment Analysis(2011)

[6]Twitter API取得方法(2019)

[7]自然言語処理(2017)

[8]Bag of wordsについて(2018)