|  |
| --- |
| **FX deep learning + NLP(Trump tweet)**  2017315051  다나카 이마노무네사토 |
|  |
| **요약**  LSTM은 다양한 시계열 예측문제 및 시간분류문제에 잘 적용되고 있고 주식, FX의 예측문제에 적당하다. 하지만, 과거의 데이터로부터 LSTM로 학습하기에는 금융의 뉴스를 고려할 수 없다는 문제가 있다.이번엔 Natural Language Processing(NLP) Sentimental을 사용하고 trump의 tweet이 가지는 FX시장의 영향을 고려한다. 정확한 시장 트렌드를 python nltk로 학습시기는 것과 과거의 데이터로부터 몇 분, 몇 시간 후의 트렌드를 LSTM로 학습시키는 것을 합친 새로운 지표를 사용하여 시장의 트렌드를 예측한다. | | |

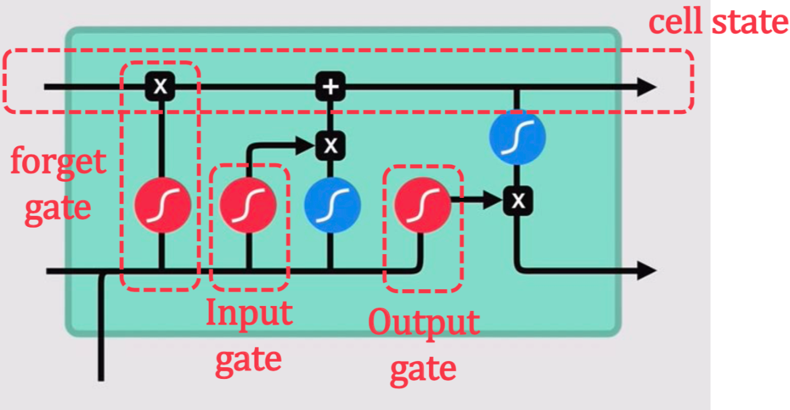
1. **프로젝트 동기 및 배경**

오늘날, 세계의 금융시장은 현대 경제에 막대한 영향을 미치고 있다. 다양한 시장에서 매일 수 조 달러의 금융 상품이 거래되고 있다. 외국외환시장(FX)도 예외가 아니다.  모든 transaction은 시간과 함께 변동하는 통가의 환율rate을 형성한다. 매일의 거래 중 고액 거래를 고려하여 이것들의 rate 변동을 예측할 수 있다면 막대한 이익을 얻을 수 있다.

**LSTM(Long Short−Term Memory)**

　Long Short-Term Memory network는 [LSTM]라 불리고 있고, 장기적인 의존관계를 학습할 수 있는 특별한 RNN이다. Hochreiter & Schmidhuber(1997)에 의해서 발표되고, 후에 연구자들이 많은 논문을 썼다. 현재도 널리 사용되고 있는 기술이다. short-term memory란 단기기억의 의미이고, 단기 기억을 장기적으로 걸쳐서 활용하는 것을 가능하게 되는 것이 LSTM의 중요한 성과이다.

RNN을 실현하기 위해서 고안됐기 때문에 이전의 정보를 잘 처리하려고 하는 목적은 같지만 안쪽 부분의 처리가 다르다.



Cell State는 상대적인 정보를 끝까지 전송한다. 학습 결과는 이곳을 지난다. Forget gate는 앞의 입력과 이번 입력이 중요한지 아닌지를 판단하여 Cell State에게 얼마만큼 전달 여부를 결정한다. Input gate는 Cell state을 얼마나 업데이트 할 것인지 결정한다. Output gate는 다음 입력에 정보를 전달한다. [1][2][3]

**Foreign Exchange(FX)**

　FX란, (Foreign Exchange)의 약자이며 정식 명칭은 외환 증거금 거래라고 한다. 다른 나라 통화의 매매하는 의미이고 ¥과＄、¥과₩、＄과€、£등 통화 자체를 사고 파는 거래다.통화의 환율은 항상 변동하고 있고, FX는 이 변동하는 환율을 사용해 이익을 낸다.

예를 들으면 ¥１＝₩１０일 때에 ¥１를 사고, 한달 후 ¥１＝₩１２가 되고 있다면 ₩２의 이익이 된다. 만약, ¥1000만을 사고 있다면 ₩200만의 이익이 된다. 하지만, 이익을 얻기 위해서는 통화가 언제 싸고 언제 비싼지 시세의 흐름을 이해하고, 결제 시기를 알 필요가 있다. 해외 정세에 좌우되는 FX의 rate는, chart을 보고 과거의 가격 변동의 경향을 파악하면서 거래를 해 나가는 것이 일반적이다.[4]



1. **기존 연구내용**

1분마다 트렌드가 오르거나 내리거나 하는지 예상하는 technical분석에 몇가지 Deep Learning학습 모델을 적용하고 사용 모델은 Convolutional Neural Network과 LSTM Neural Network였다. [1]

1. **문제정의**

기존 연구 내용으로 설명한 연구에서는 technical분석(데이터 해석)만 사용되고 있다. 즉 과거의 FX 거래 데이터만을 보고 트렌드를 판단하고 있다. 그러나, 실제로 트렌드의 상하를 판단하려면 금융 뉴스도 고려하여 트렌드 예상해야 한다. 금융뉴스 기사에는 금융상품의 가격 또는 환율에 영향을 미치는 중요한 정보가 포함되어 있는 것이 많지만, 이번에 정보는 고려되어 있지 않다. 즉, fundamental분석(경제적 지표)을 사용되지 않은 것이 문제이다.

1. **연구목표**

본 연구에서는 문제 정의를 고려하여 technical분석(데이터 해석)에 덧붙여 fundamental분석(경제적인 뉴스, 요인의 발언)을 고려하여 자연 언어 학습을 시킨다. 다른 FX예측문제에서는 NLP을 고려하고 있는 것이 없음으로, 이번에 처음으로 LSTM과 NLP을 합친 새로운 모델을 만들어 가기로 한다.

1. **선행연구**

**Technical분석**

technical분석이란 과거의 시세의 움직임으로부터, 장래의 가격변동을 예상하는 분석 방법이다. 과거의 가격변동을 그래프로 한 chart 등을 사용하는 경우가 많다. 지표에서는 Candlestick char, support line and resistance line, Fibonacci retracement, Moving average line、RSI、MACD、Stochastic, Bollinger Band, Dow theory etc.

**이동 평균(moving average)**

어느 일정기간의 가격(보통 close)을 평균으로 묶은 선.

기본은 단기선, 중기선, 장기선 3종류로 나눌 수 있다.

일반적으로 단기선은５일, ６일, １２일.  중기선은２０일, ２５일, ５０일. 장기선１００일, １４４일, ２００일, ２３３일이 사용된다. 이런 평균선을 봐야 trend을 파악할 수 있다.[4]

**기존 Deep Learning LSTM 모델 기계 학습**

LSTM 모델에 맞게 데이터를 전처리를 한다. LSTM의 연속되는 값은 이동 평균 기간에 맞게 결정한다. 논문에서는 단기 이동 평균 선의 5일을 사용하고 있다. 우선 Train데이터를 5일 동안 하나로 정리한다. 각 5일 이동 평균을 구한다. Target이 되는 것은 전의 이동 평균에서 뒤의 이동 평균을 뺄 때 값이 0보다 크면 전의 기간보다 값이 올랐다고 간주하고 +1로 한다. 값이 0보다 작은 경우 앞의 기간보다 값이 하강했다고 간주하고 -1로 한다. 사용한 parameter는 Activation Function(relu, softmax)、Opimizer(Adam)、Loss function(Categorical CrossEntropy)이다.[1]

**Fundamental분석**

fundamental분석이란 환율에 관한 기초적인 정보를 바탕으로 시세를 분석하는 것이다. 경제지표나 금리, 그 외 통화나 나라에 관한 여러가지 요인이 분석 재료가 되고 있다. 요인에는 경제지표 외에도 중요 인물들의 발언도 영향을 받는다. 예를 들어 FRB（Federal　Reserve　Board）의장, ECB（European Central Bank）총재, 일본은행총재의 발언이 주목된다. 그 중에서도 SNS을 통해 Donald Trump의 tweet가 FX시장에 영향을 주고 있다.

2018년 7월 23일 로이터의 조사에서 Trump와 FX의 관계성에 관한 기사가 있었다.[4]



**자연언어 전처리**

전 처리의 대략적인 흐름으로서 cleaning처리→문장의 단어분할→단어 정규화→ stop word의 제거로 대략적으로 흐른다. 먼저 cleaning처리에서는, ~~<h1>~~FX news usd jpy~~</h1>~~처럼 text안의 노이즈를 제거하는 작업을 한다.

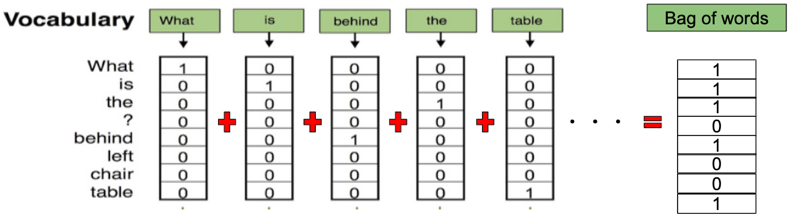
문장의 단어분할에서는 원문「I will go to sea Today」라는 문장을「I/will/go/to/sea/Today처럼 분할시킨다. 단어 정규화에서는Apple→apple처럼 문장의 동일 대문자 소문자 변환을 한다. Stop word제거에서는 태스크를 풀기 위해 도움이 되지 않는다고 생각되는 단어를 제거한다. [7]

**Bag of Words**

Bag of Words라는 것은 자연언어 처리에 있어서 자연언어(인간이 일상에서 사용하고 있는 언어)로 기술된 데이터, 즉 글을 벡터로 표현하는 방법이다.

Bag of Words는 다음의 3단계에서 만들 수 있다.[8]

1. 수치변환
2. one hot vector
3. 더하기



1. **문제해결방안**

선행연구를 통해 문제 정의로 제시하고 있는 자연언어 처리를 맞춘 LSTM의 FX 트렌드 예상을 실현하기 위해 이하의 해결방법을 정의하기로 했다.

1.단기, 중기, 장기이동평균을 LSTM 현식으로 데이터 전처리를 한다.

2. 웹에서 FX에 관한 뉴스를 추출하고 자연언어처리를 한다.

3.(2) 에서 FX에 관한 중요한 단어를 취득한다.

4. Donald Trump의 Tweet을 API에서 취득한 그 중 (2)에서 취득한 FX용어가 포함된 Tweet을 선택하고 자연 언어 처리를 한다. Tweet이 긍정적인지 부정적인지를 알아보고 기계학습의 타겟으로 삼는다. [5][6]

5. 실제로 Deep Learning LSTM모델을 사용하고 기계학습을 한다.

6. 일반적인 Deep Learning LSTM모델과 이번의 모델과 비교를 한다.

단기, 중기, 장기 이동 평균을 LSTM의 형식에 데이터 전처리를 하는 것은 이번 단기 이동 평균을 사용하는 실제 데이터의 shape은(Total FX data, 5 days, columns number)형식이다. Web에서 FX에 관한 뉴스를 추출하여 자연언어 처리를 한다면, google에서 USD/JPY에 관한 기사를 web scraping하여 취득하고, 자연언어 전처리를 했다. 다음으로 Bag of words을 사용하고 그 뉴스에서 자주 쓰이는 명사 단어 리스트를 작성한다.Trump tweet을 전처리하고 명사 단어가 포함된 단어가 있는 tweet만을 골라 감정 분석을 한다. 긍정적이면 +1부정적이라면-1로 target로 한다. 이를 LSTM 기계학습 모델을 사용하여 학습시키고 기존 모델과 비교를 한다.

데이터 나눈 방법

Twitter API뿐만 아니라Kaggle도 사용한다.

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)

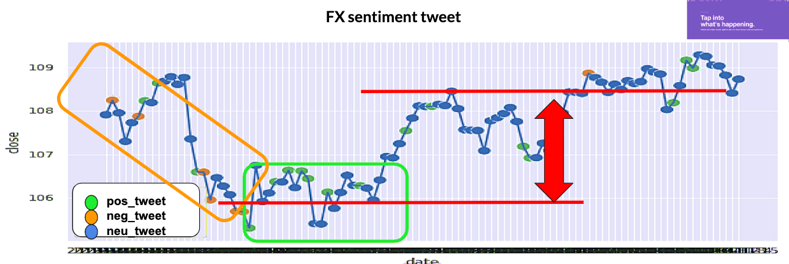
Date : 2015/07/17 ~ 2016/11/11

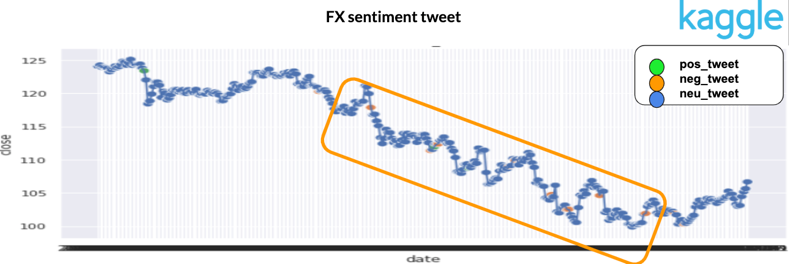
Train data : 2015/07/17 ~ 2015/08/05

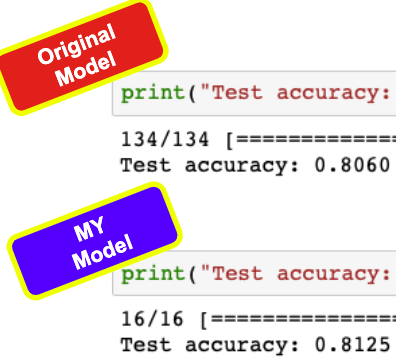
Test data: 2015/08/06 ~ 2016/11/11

1. **결과・평가**

　초록색이 긍정적 Tweet, 오렌지색이 부정적 Tweet, 파란색이 중립인 Tweet이다. 실제로 긍정적, 부정적으로 나누어 가시화해 본 결과, 부정적인 Tweet이 모여 있는 곳에서는 트렌드가 내려가는 경향을 볼 수 있었고, 긍정적인 Tweet이 모이면 나중에 트렌드가 상승하는 힘이 작용한다는 것을 알 수 있다. 이를 통해 트럼프 대통령의 Tweet이 적지 않게 FX에 영향이 있다는 것을 알게 되었다.

LSTM 기계학습도 시키고, 기존 모델과 이번 모델의 비교를 한 결과, 조금이지만 이번 모델이 정확성이 더 높았다는 것을 알 수 있다.





1. **Discussion**

　우선 무엇보다도 아쉬운 것은 Twitter API의 제한에서 2019/07/15~2019/11/15까지 충분한 Tweet수를 취득하지 못한 것이다. 하지만 이번 연구는 LSTM과 NLP을 합친 FX의 연구가 언급된 선행 연구의 사례가 없었기 때문에, 새로운 지표를 사용해 연구했다. 그 결과 LSTM에서 FX 트렌드 예상과 NLP 학습에 따라 경제적 지표도 조금만 고려해볼 수 있었다. Twitter 제한 문제가 해결되면 더 좋은 연구결과가 나올 것이다.

1. **연구의 중요성**

　이번 연구에서 Donald Trump의 Tweet이 FX의 환율과 관련되어 있다는 것을 알았다. 그리고 과거 움직임의 패턴과 경제적인 지표를 고려해 FX 트렌드를 깊게 알 수 있었다. 이 연구가 최종적으로 이익을 얻을 수 있도록 하기 위한 참고가 된다면 좋겠다고 생각한다.

1. **향후 발전 계획**

　Web scraping 기술을 구사해 경제적인 지표의 데이터를 대량으로 모은다. 또 다른 연구자들이 이 연구를 참고하여 연구 진행에 도움이 되기를 바란다.

1. **참고문헌**

[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)